INNER EXPERIENCES: THEORY, MEASUREMENT, FREQUENCY, CONTENT, AND FUNCTIONS

EDITED BY : Alain Morin, Thomas M. Brinthaupt and Jason D. Runyan
PUBLISHED IN : Frontiers in Psychology

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

INNER EXPERIENCES: THEORY, MEASUREMENT, FREQUENCY, CONTENT, AND FUNCTIONS

Topic Editors:
Alain Morin, Mount Royal University, Canada
Thomas M. Brinthaupt, Middle Tennessee State University, USA
Jason D. Runyan, Indiana Wesleyan University, USA

Image: "Topiary Trip" by Famira Racy, © 2015 cc 3.0

One fundamental topic of scientific inquiry in psychology is the study of what William James called the "stream of consciousness", our ongoing experience of the world and ourselves from within—our inner experiences. These internal states (aka "stimulus-independent thoughts") include inner speech, mental imagery, feelings, sensory awareness, internally produced sounds or music, unsymbolized thinking, and mentalizing (thinking about others' mental states). They may occur automatically during mind-wandering (daydreaming) and resting-state episodes, and may focus on one's past, present, or future ("mental time travel"--e.g., autonoetic consciousness). Inner experiences also may take the form of intrusive or ruminative thoughts.

The types, characteristics, frequency, content, and functions of inner experiences have been studied using a variety of traditional methods, among which questionnaires, thought listing procedures (i.e., open-ended self-reports), thinking aloud techniques, and daily dairies. Another approach, articulatory suppression, consists in blocking participants' use of verbal thinking while completing a given task; deficits indicate that inner speech plays a causal role in normal task completion. Various thought sampling approaches have also been developed in an effort to gather more ecologically valid data. Previous thought sampling studies have relied on beepers that signal participants to report aspects of their inner experiences at random intervals. More recent studies are exploiting smartphone technology to easily and reliably probe randomly occurring inner experiences in large samples of participants.

These various measures have allowed researchers to learn some fundamental facts about inner experiences. To illustrate, it is becoming increasingly clear that prospection (future-oriented thinking) greatly depends on access to autobiographical memory (past-oriented thinking), where recollection of past scenes is used as a template to formulate plausible future scenarios.

The main goal of the present Research Topic was to offer a scientific platform for the dissemination of current high-quality research pertaining to inner experiences. Although data on all forms of inner experiences were welcome, reports on recent advances in inner speech research were particularly encouraged. Here are some examples of topics of interest: (1) description and validation of new scales, inventories, questionnaires measuring any form of inner experience; (2) novel uses or improvements of existing measures of inner experiences; (3) development of new smartphone technology facilitating or broadening the use of cell phones to sample inner experiences; (4) frequency, content, and functions of various inner experience; (5) correlations between personality or cognitive variables and any aspects of inner experiences; (6) philosophical or theoretical considerations pertaining to inner experiences; and (7) inner experience changes with age.

Table of Contents

frontiers
in Psychology

Editorial: Inner Experiences: Theory, Measurement, Frequency, Content, and Functions

Alain Morin[1]*, Jason D. Runyan[2] and Thomas M. Brinthaupt[3]

[1] Department of Psychology, Mount Royal University, Calgary, AB, Canada, [2] Department of Psychology, Indiana Wesleyan University, Marion, IN, USA, [3] Department of Psychology, Middle Tennessee State University, Murfreesboro, TN, USA

It is safe to posit that human beings have been interested in their own inner mental experiences from the moment they became aware of them, arguably over 60,000 years ago (Leary, 2004). In sharp contrast, growth in the actual scientific examination of these inner experiences is remarkably recent (e.g., Csikszentmihalyi and Figurski, 1982; Klinger and Cox, 1987–1988; Goldstein and Kenen, 1988; Hurlburt, 1990). Inner speech, in particular, has been the focus of even more recent efforts (e.g., Morin et al., 2011; Brinthaupt and Dove, 2012; Hurlburt et al., 2013; Alderson-Day and Fernyhough, 2015; Alderson-Day et al., 2015). We present 14 articles that cover theoretical ideas as well as current research results pertaining to the measurement, frequency, content, and functions of inner experiences. In what follows we summarize some exciting key findings highlighted in this research topic.

CONTENT AND FUNCTIONS OF INNER EXPERIENCES/INNER SPEECH

There are large individual differences in inner experiences (i.e., inner speech, inner seeing, unsymbolized thinking, feelings, sensory awareness). In particular, resting states (relaxing without falling asleep with eyes open) seem to differ substantially from one participant to the next (Hurlburt et al., 2015). The resting state includes several distinct dimensions such as thinking about others' mental states, planning, sleepiness, bodily awareness, inner speech, mental imagery, and health concerns (Diaz et al., 2014). Inner speech probably represents a speaking activity that does not have a proper function in cognition. Rather, it inherits the array of functions of outer speech (Martínez-Manrique and Vicente, 2015). Furthermore, the relation between inner and outer speech is more complex than initially thought—e.g., patients with inner speech deficits can still overtly name objects (Langland-Hassan et al., 2015).

MEASUREMENT OF INNER EXPERIENCES

Many existing self-consciousness scales measure various related—yet different—self-reflective constructs such as adaptive-maladaptive, public-private self-consciousness, and mindfulness (DaSilveira et al., 2015). Smartphone technology allows us to significantly refine current thought sampling methods (e.g., ecological momentary assessment)—e.g., by gathering repeated sampling within various situations of daily life in very large samples, and allowing the capture of dispositional expressions (Runyan and Steinke, 2015). Repeated sampling can also promote self-awareness and

mindfulness. Also, inner speech measured with self-report questionnaires and thought sampling procedures poorly correlate, suggesting that self-report approaches may not be valid (Alderson-Day and Fernyhough, 2015; also see Uttl et al., 2011). One exception is the Self-Talk Scale, which exhibits good psychometric qualities in multiple studies (Brinthaupt et al., 2015).

MEMORY, TIME PERCEPTION, AESTHETIC EXPERIENCE, AND IMAGINATION

When compared to younger participants, older adults report more details (e.g., when and where of events, emotions experienced, people and objects involved) in their remote/recent autobiographical memories (Gardner et al., 2015). Delayed video presentations of one's own body image alters time perception (Fritz et al., 2015). Aesthetic insight (i.e., the "aha" phenomenon) occurs as artistic material becomes more complex and determinate, increases liking, and is preceded by increased interest, supporting the theory that interest is increased by the expectation of understanding (Muth et al., 2015). Imagination and creative thought likely emerge from the conceptual and factual knowledge accumulated throughout our lives—our semantic learning (Abraham and Bubic, 2015).

Based on the 14 articles presented here, we suggest that future work on inner experiences expand on (a) future-oriented thinking, (b) naturally occurring mentalizing, (c) mindfulness, (d) abnormal manifestations of inner experiences, (e) correlations between inner experiences and actual behavior, (f) inner experience changes with age, (g) inner experiences in altered states of mind and religious states, as well as (h) in infants and non-human animals.

AUTHOR CONTRIBUTIONS

AM wrote the editorial; JR and TB edited it.

REFERENCES

Abraham, A., and Bubic, A. (2015). Semantic memory as the root of imagination. *Front. Psychol.* 6:325. doi: 10.3389/fpsyg.2015.00325

Alderson-Day, B., and Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychol. Bullet.* 141, 931–965. doi: 10.1037/bul0000021

Alderson-Day, B., Weis, S., McCarthy-Jones, S., Moseley, P., Smailes, D., and Fernyhough, C. (2015). The brain's conversation with itself: neural substrates of dialogic inner speech. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsv094. [Epub ahead of print].

Brinthaupt, T. M., Benson, S. A., Kang, M., and Moore, Z. D. (2015). Assessing the accuracy of self-reported self-talk. *Front. Psychol.* 6:570. doi: 10.3389/fpsyg.2015.00570

Brinthaupt, T. M., and Dove, C. T. (2012). Differences in self-talk frequency as a function of age, only-child, and imaginary childhood companion status. *J. Res. Person.* 46, 326–333. doi: 10.1016/j.bbr.2011.03.031

Csikszentmihalyi, M., and Figurski, T. (1982). Self-awareness and aversive experience in everyday life. *J. Person.* 50, 14–26. doi: 10.1111/j.1467-6494.1982.tb00742.x

DaSilveira, A., DeSouza, M. L., and Gomes, W. B. (2015). Self-consciousness concept and assessment in self-report measures. *Front. Psychol.* 6:930. doi: 10.3389/fpsyg.2015.00930

Diaz, B. A., Van Der Sluis, S., Benjamins, J. S., Stoffers, D., Hardstone, R., Mansvelder, H. D., et al. (2014). The ARSQ 2.0 reveals age and personality effects on mind-wandering experiences. *Front. Psychol.* 5:271. doi: 10.3389/fpsyg.2014.00271

Fritz, T. H., Steixner, A., Boettger, J., and Villringer, A. (2015). Losing track of time through delayed body representations. *Front. Psychol.* 6:405. doi: 10.3389/fpsyg.2015.00405

Gardner, R. S., Mainetti, M., and Ascoli, G. A. (2015). Older adults report moderately more detailed autobiographical memories. *Front. Psychol.* 6:631. doi: 10.3389/fpsyg.2015.00631

Goldstein, G., and Kenen, R. (1988). "Internal dialogue" in a "normal" population: the implications for health promotion. *Health Promot. Int.* 3, 249–257.

Hurlburt, R. T. (1990). *Sampling Normal and Schizophrenic Inner Experience*. New York, NY: Plenum Press.

Hurlburt, R. T., Alderson-Day, B., Fernyhough, C., and Kühn, S. (2015). What goes on in the resting-state? A qualitative glimpse into resting-state experience in the scanner. *Front. Psychol.* 6:1535. doi: 10.3389/fpsyg.2015.01535

Hurlburt, R. T., Heavey, C. L., and Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Conscious. Cogn.* 22, 1477–1494. doi: 10.1016/j.concog.2013.10.003

Klinger, E., and Cox, W. M. (1987–1988). Dimensions of thought flow in everyday life. *Imaginat. Cognit. Person.* 7, 105–128.

Langland-Hassan, P., Faries, F. R., Richardson, M. J., and Dietz, A. (2015). Inner speech deficits in people with aphasia. *Front. Psychol.* 6:528. doi: 10.3389/fpsyg.2015.00528

Leary, M. R. (2004). *The Curse of the Self: Self-Awareness, Egoism, and the Quality of Human Life*. New York, NY: Oxford University Press.

Martínez-Manrique, F., and Vicente, A. (2015). The activity view of inner speech. *Front. Psychol.* 6:232. doi: 10.3389/fpsyg.2015.00232

Morin, A., Uttl, B., and Hamper, B. (2011). Self-reported frequency, content, and functions of inner speech. *Proc. Soc. Behav. J.* 30, 1714–1718. doi: 10.1016/j.sbspro.2011.10.331

Muth, C., Raab, M. H., and Carbon, C.-C. (2015). The stream of experience when watching artistic movies. Dynamic aesthetic effects revealed by the Continuous Evaluation Procedure (CEP). *Front. Psychol.* 6:365. doi: 10.3389/fpsyg.2015.00365

Runyan, J. D., and Steinke, E. G. (2015). Virtues, ecological momentary assessment/intervention and smartphone technology. *Front. Psychol.* 6:481. doi: 10.3389/fpsyg.2015.00481

Uttl, B., Morin, A., and Hamper, B. (2011). Are self-report questionnaires of inner speech valid and reliable? *Proc. Soc. Behav. J.* 30, 1719–1723. doi: 10.1016/j.sbspro.2011.10.332

What goes on in the resting-state? A qualitative glimpse into resting-state experience in the scanner

Russell T. Hurlburt[1]*, Ben Alderson-Day[2], Charles Fernyhough[2]* and Simone Kühn[3,4]

[1] *Department of Psychology, University of Nevada, Las Vegas, Las Vegas, NV, USA,* [2] *Department of Psychology, Durham University, Durham, UK,* [3] *Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany,* [4] *Clinic and Polyclinic for Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf, Hamburg, Germany*

The brain's resting-state has attracted considerable interest in recent years, but currently little is known either about typical experience during the resting-state or about whether there are inter-individual differences in resting-state phenomenology. We used descriptive experience sampling (DES) in an attempt to apprehend high fidelity glimpses of the inner experience of five participants in an extended fMRI study. Results showed that the inner experiences and the neural activation patterns (as quantified by amplitude of low frequency fluctuations analysis) of the five participants were largely consistent across time, suggesting that our extended-duration scanner sessions were broadly similar to typical resting-state sessions. However, there were very large individual differences in inner phenomena, suggesting that the resting-state itself may differ substantially from one participant to the next. We describe these individual differences in experiential characteristics and display some typical moments of resting-state experience. We also show that retrospective characterizations of phenomena can often be very different from moment-by-moment reports. We discuss implications for the assessment of inner experience in neuroimaging studies more generally, concluding that it may be possible to use fMRI to investigate neural correlates of phenomena apprehended in high fidelity.

Keywords: resting state, descriptive experience sampling (DES), fMRI, default mode network (DMN), Resting State Questionnaire (ReSQ), mind wandering

Introduction

In resting-state functional magnetic resonance imaging (rfMRI), spontaneous changes in the blood oxygen dependent (BOLD) signal can be used to study networks of brain areas that are functionally connected and tend to co-activate when a participant is not performing any explicit task, that is, when a participant is in what is often referred to as the "resting state." These studies produce activity in a consistent network of brain regions, including lower precuneus, superior and inferior anterior medial frontal regions, and posterior lateral parietal cortices (Gusnard and Raichle, 2001). Initially, these regions were identified because they were found to be consistently deactivated when tasks are performed. The consistency with which this set of brain regions decrease in activity during tasks and increase during fixation or resting has led to the notion of a so-called "default mode" network of the brain (Buckner et al., 2008).

Scientists interested in rfMRI have provided a wide variety of characterizations of the kinds of experiences and processes that are ongoing when default-mode brain regions are active, as exemplified in **Table 1**. As those characterizations demonstrate, there is wide variability in the descriptions of phenomena in the resting state. Researchers are increasingly sensitive to the potential phenomenological heterogeneity of subjective experience in the resting state (Smallwood and Schooler, 2015), denoted here as 'mind wandering' in accord with popular usage (Callard et al., 2013). For example, Gorgolewski et al. (2014) demonstrated relations between the content and form of self-generated thoughts in the resting state (such as imagery and future-related thinking) and specific intrinsic neural activity patterns. Ruby et al. (2013) showed that the relation between the emotional content of thoughts and subsequent mood was modulated by the socio-temporal content of the thoughts, specifically their relatedness to the past. At the same time, there is a growing recognition that psychological experiences collected under the umbrella of mind wandering are not necessarily the experiential manifestation of neural activity in the resting state, and that the relation between experiential phenomena and neural state is complex (Raichle, 2011; Fox and Christoff, 2014).

Most characterizations of mind wandering as the psychological counterpart of the brain's resting state follow from indirect theoretical considerations (researchers set tasks for participants in the scanner and then theorize about what is *not* ongoing in those tasks; Callard and Margulies, 2014) or retrospective characterizations. Researchers have recently developed three questionnaires that ask participants retrospectively to characterize their resting state cognition: the Resting State Questionnaire (ReSQ; Delamillieure et al., 2010), the Amsterdam Resting State Questionnaire (ARSQ; Diaz et al., 2013, and its revised and extended counterpart the ARSQ 2.0 Diaz et al., 2014), and the New York Cognition Questionnaire (NYC-Q; Gorgolewski et al., 2014). These questionnaires typically ask volunteers to participate in an fMRI resting-state session (or online analog thereof; Diaz et al., 2014), exit the scanner, and then immediately characterize their in-scanner resting state experience. The ReSQ asks participants to use visual-analog scales to estimate the proportion of their resting-state time that they had been engaged in visual imagery, in inner language, in somatosensory awareness, in inner musical experience, and in the mental manipulation of numbers. The ARSQ 2.0 uses 30 Likert-scale items divided into ten factors (discontinuity of mind, theory of mind, self, planning, sleepiness, comfort, somatic awareness, health concern, visual thought, and verbal thought). The NYC-Q asks participants to report on the content and form of their self-generated thoughts using Likert scales. Factor analysis of the NYC-Q has revealed five main content factors of resting-state experiences: past, future, positive, negative, and social experiences, and three main form factors: words, images, and thought specificity. All three of these questionnaires ask respondents to characterize their resting state in general and therefore do not characterize particular moments. However, Hurlburt and Heavey (2015) argued that retrospective questionnaires may not be adequate to characterize moments of inner experience. They held that retrospective reports about inner experience are perhaps more influenced by the participant's presuppositions about experience than by the participant's experience itself, that retrospections are skewed by reporting biases such as recency or salience, and so on.

There have been a few studies that have sought to overcome the retrospectiveness limitations by using experience sampling to examine experiences at specific moments during the resting state. For example, Christoff et al. (2009) had subjects in an fMRI scanner perform a boring go/no-go task and intermittently presented thought probes. Each probe asked participants to report on their mental state using two Likert scales: one asked whether attention was focused on the task (rated from *completely*

TABLE 1 | Characterizations of experiences when the default mode is active.

	Characterization	Source
1	"unconstrained verbally mediated thoughts"	Shulman et al., 1997, p. 648
2	"semantic knowledge retrieval, representation in awareness, and directed manipulation of represented knowledge for organization, problem-solving, and planning"	Binder et al., 1999, p. 80
3	"active retrieval of past experiences and planning of future experiences"	Andreasen et al., 1995, p. 1576
4	"retrieval and manipulation of past events, both personal and general, in an effort to solve problems and develop future plans"	Greicius et al., 2003, p. 257
5	"enhanced watchfulness toward the external environment (e.g., waiting for upcoming task-relevant stimuli or attending to scanner noise and incidental light)"	Gilbert et al., 2007, p. 43
6	"inner thought, self-reflective thinking in terms of planning for the future, or simulation of behavior… interrupted… into a… extrospective… state of mind… characterized… as increased attention and readiness,… sensorimotor planning for future routes of action in response to potential changes in the inner and outer environment"	Fransson, 2005, p. 26
7	"not focused on the external environment,… internally focused tasks including autobiographical memory retrieval, envisioning the future, and conceiving the perspectives of others"	Buckner et al., 2008, p. 1
8	"spontaneous, internally directed cognitive processes"	Andrews-Hanna et al., 2010, p. 322
9	"spontaneous mental contents which are unrelated to perception and coordinate[d]… so that they are maintained in the face of competing sensory information"	Smallwood et al., 2012, p. 67
10	"an ultimate state of inspection of the self"	Wicker et al., 2003, p. 229
11	"stable, unified perspective of the organism relative to its environment (a 'self')"	Gusnard and Raichle, 2001, p. 692

---

on task to completely off task); the second asked whether the subject was aware of where their attention was focused (rated from completely aware to completely unaware). However, such studies have focused on rating one or two aspects of experience and have not tried to provide descriptions of actual ongoing experience.

Thus there are to date no investigations that have sought to provide high fidelity descriptions of the phenomena that are ongoing in the resting state. Understanding the experiential details of the resting state is important because the resting state is typically the baseline against which the results of particular tasks are compared, as well as being an important target of investigation in its own right. Hurlburt and Heavey (2015) suggested that descriptive experience sampling (DES; Hurlburt, 1990, 1993, 2011a; Hurlburt and Akhter, 2006; Hurlburt and Heavey, 2006; Hurlburt and Schwitzgebel, 2007) may be capable of producing high fidelity descriptions of experience. In its typical application, DES uses a random beeper to interrupt participants in their natural environments. Participants are to attend to the experience that was ongoing at the moment of the onset of the beep and to jot down notes about that experience. A typical participant receives six such beeps in typically a 3-h window. Later that day (or the next day), the participant meets with the investigator in what DES calls an expositional interview designed to discover the details of the six experiences and "iteratively" to improve the quality of subsequent sampling. The sample/interview procedure is then repeated over a number (typically 4–6) of sampling days.

Descriptive experience sampling has been compared to and contrasted with a variety of qualitative and related methods (see **Table 2**). In broad strokes, DES differs from other methods in that DES aims at pristine inner experience (Hurlburt, 2011a), experience that is directly apprehendable at a moment; its view that people often do *not* know the characteristics of their own experience unless trained to apprehend them, probably requiring an iterative method (Hurlburt, 2009, 2011a); its minimization of retrospection; and its methods of bracketing of presuppositions (Hurlburt and Schwitzgebel, 2007, 2011; Hurlburt, 2011a).

We have seen, then, that (a) it would be desirable to understand the phenomenology of experience while in the resting state in the scanner; and that (b) DES, with its iterative training, may provide high fidelity access to the phenomenology of experience in a way that can be integrated with fMRI (Kühn et al., 2014). The present study seeks to combine those two considerations, a non-trivial exercise because the duration of DES studies (typically measured in days) is far greater than is typically available in the magnetic resonance imaging (MRI) resting state sessions (typically measured in minutes). As illustrated in **Figure 1**, the present study attempted to overcome that obstacle first by preliminarily training participants in DES in their natural environments prior to involvement in the scanner and then by using DES in extended-duration fMRI resting state sessions. The extended-duration resting-state sessions involved nine 25-min fMRI sessions for each participant ($9 \times 25 = 225$ min of resting state per participant instead of the more usual 5 or 10 min). Participants were

TABLE 2 | Comparing DES with qualitative and related methods.

Method	Comparison reference
Kvales' qualitative research interview	Hurlburt and Heavey, 2006, Chap. 12
Giorgi's phenomenological psychology	Hurlburt and Heavey, 2006, Chap. 12
Stern's micro-analytic interview	Hurlburt, 2011a, Chap. 7
Vermersch's explicitation interview	Hurlburt, 2011b
Petitmengin's second-person interview	Hurlburt and Akhter, 2006
van Manen's hermeneutic phenomenological inquiry	Heavey et al., 2010
Moustakas's human science research	Heavey et al., 2010
Armchair introspection	Hurlburt and Schwitzgebel, 2007, Chap. 11; 2011
Eyewitness testimony	Hurlburt and Schwitzgebel, 2007, Chap. 11
Questionnaires	Hurlburt and Heavey, 2015
Non-DES experience sampling methods	Hurlburt and Heavey, 2015

given the same instructions as are usual in resting-state studies.

Materials and Methods

Participants

Five native English-speaking (because RTH would be the lead interviewer) participants who currently lived in Berlin (because MRI scanning would take place at the Max Planck Institut für Bildungsforschung) participated on the basis of informed consent and with ethical committee approval according to the Declaration of Helsinki. All participants had normal or corrected-to-normal vision. No participant had a history of neurological, major medical, or psychiatric disorder. The participants (three females, two males) had a mean age of 22.4 (ranging from 18 to 30) and all but one (male) were right-handed.

Measures

The ReSQ (Delamillieure et al., 2010) is a semi-structured questionnaire that asks participants to characterize their inner experiences when they had been resting quietly in a MRI scanner.

Descriptive experience sampling was performed as described in Hurlburt (2011a) and elsewhere. DES is primarily an idiographic procedure, allowing and encouraging the examination of phenomena that may be idiosyncratic to particular individuals, but it has identified five phenomena that are characteristic of many individuals, which we will call the 5FP (five frequent phenomena): inner speaking (the experience of speaking to oneself in one's own voice but without any external sound or motor movement; Hurlburt et al., 2013); inner seeing (experiencing imaginary seeing); unsymbolized thinking (the directly apprehended experience of thinking that is not accompanied by the apprehension of words, visual images, or any other symbols; Hurlburt and Akhter, 2008a,b); sensory awareness (experience where a particular focus is on
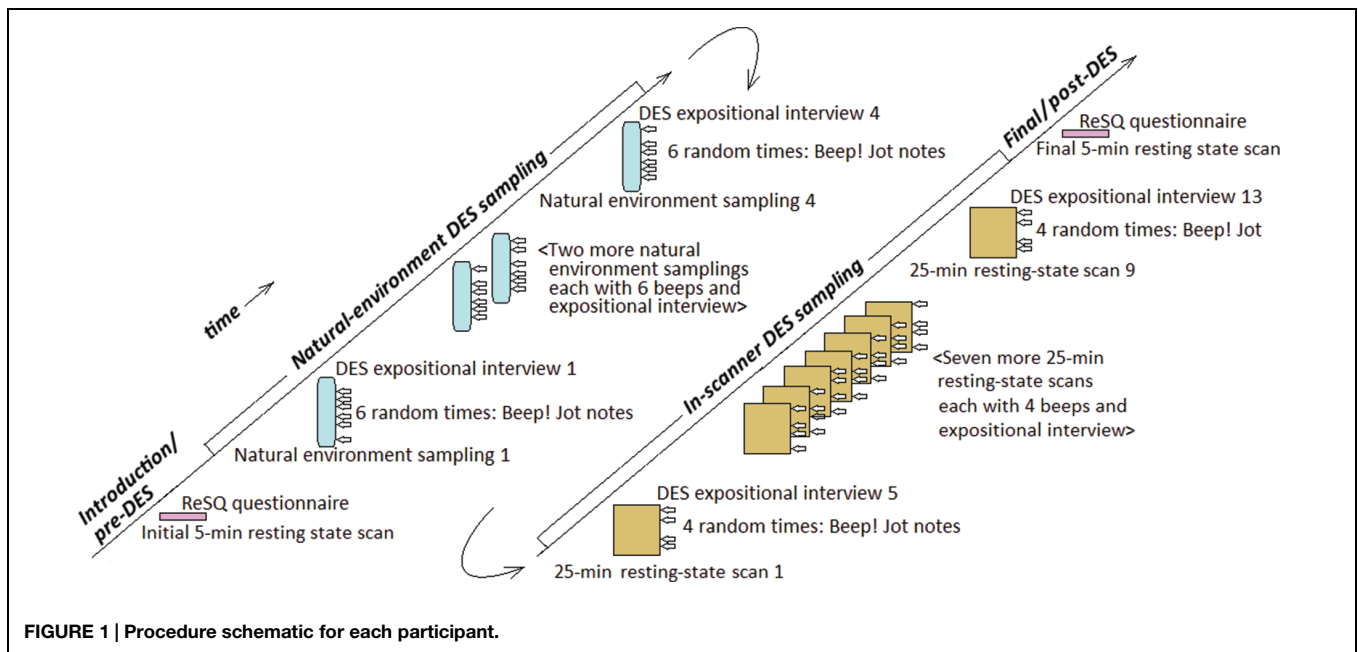
FIGURE 1 | Procedure schematic for each participant.

a sensation, not for any instrumental utility; Hurlburt et al., 2009); and feelings (the experience of emotion; Heavey et al., 2012).

Other measures not relevant here were administered as part of a larger study.

Procedure

A bird's-eye view of the procedure for each participant is illustrated in **Figure 1**. Each participant was scheduled for 19 sessions, generally across a 2-weeks period, which were divided into four phases. Schedules were individualized for each participant; for example, scanner sessions were generally twice a day, but because of holiday or other pressures were occasionally once or three times on some days. DES instructions were also individualized; a tenet of DES is to be candid, and participants with more questions got more initial instructions.

In the *introduction/pre-DES resting state* phase (**Figure 1**, bottom left), we fully explained the study, administered initial questionnaires not relevant to the present report, and familiarized the participant with the MRI scanner and procedures. Then the participant entered the scanner, where we conducted a structural scan and then a 5-min resting state scan according to standard procedures. The resting-state instructions were "please close your eyes and relax, without falling asleep." Immediately following the resting-state scan, the participant exited the scanner and filled out the ReSQ questionnaire under supervision of a psychologist.

In the *natural-environment DES sampling* phase (**Figure 1**, middle left), which began typically immediately after the completion of the ReSQ, we instructed the participant in the use of the DES beeper and the sampling task (Hurlburt, 2011a; Hurlburt et al., 2013): the participant was to wear the beeper in the participant's natural environment for ~3 h, during which the participant would hear (through an earphone) six randomly occurring beeps. Immediately after each beep the participant

was to jot down notes (in a supplied small notebook) about the ongoing inner experience—the experience that was "in flight" at the moment the beep sounded. Following the DES instructions, the participant proceeded to the natural environment, wore the DES beeper, and, when beeped, collected six experience samples. Later that day or the next day the participant returned for the first DES expositional interview about those six beeped experiences; this interview was conducted by RTH and at least one and as many as four additional interviewers (the study was part of a training program), usually including SK and sometimes CF or BA-D. The expositional interview (following the DES procedure) was "iterative" (Hurlburt, 2009, 2011a), designed to increase, across sampling days, the participant's skill in apprehending and describing inner experience. Following this interview, the participant returned to his or her everyday environment and responded to six more random beeps, again jotting notes about the ongoing experiences. The following day the participant returned for a second expositional interview about the second-sampling-day's six beeped experiences. This sequence was repeated twice more, so that the participant sampled in a total of four natural-environment periods, each followed by an (iterative) expositional interview. Theoretically, this procedure could produce $4 \times 6 = 24$ natural-environment experience samples. We (as is typical in DES) considered the first day's samples unreliable and discarded them from subsequent analysis. The remaining days often produce less than the maximum six samples because the 1-h expositional interview runs out of time before all samples are discussed. Undiscussed samples are discarded. Hurlburt (2011a) has found that four such iterative sampling day/expositional interviews typically result in skill acquisition adequate for the remainder of this study; however, more interviews could be scheduled if the expositional interviews suggested it; one participant (#3) had one additional sampling day to ensure we understood each other about what was or was

not inner speaking. This procedure resulted in a variable number of natural-environment samples, ranging from 13 to 22.

In the *in-scanner DES sampling* phase (**Figure 1**, middle right), the participant (having been trained in DES in the natural environment) entered the scanner for a 25-min session with resting-state instructions "please relax, without falling asleep and do keep your eyes open." At four quasi-random times, the participant received a DES beep through a headphone (just as in the natural environment except the in-scanner beep automatically terminated after 1.5 s, whereas the natural environment beeps must be terminated by the participant). Immediately after each beep, the participant jotted a few notes about the ongoing experience on a clipboard positioned on the lap (viewable through a mirror). Immediately after exiting the scanner, the participant participated in a DES expositional interview about the four randomly beeped experiences. This in-scanner sequence (25-min fMRI scan/four beeps with jotted notes/expositional interview) was repeated eight more times, typically spread over five days, resulting in $9 \times 4 = 36$ random samples of experience occurring in $9 \times 25 = 225$ min of fMRI scanning for each participant.

In the *final/post DES* phase (**Figure 1**, upper right), the participant entered the scanner for another structural scan and a final 5-min standard resting-state scan using the same instructions and procedures as in the first 5-min resting state scan. The participant then filled out the ReSQ questionnaire under supervision of a psychologist and then was candidly debriefed.

DES Quantitative Analysis
RTH and at least one additional person present at the interview independently judged whether each of the 5FP was present at each sample; discrepancies were resolved by discussion.

Scanning Procedure
Images were collected on a 3T Magnetom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany) using a 32-channel radio frequency head coil. Structural images were obtained using a three-dimensional T1-weighted magnetization-prepared gradient-echo sequence (MPRAGE) based on the ADNI protocol[1] (repetition time [TR] = 2500 ms; echo time [TE] = 4.77 ms; TI = 1100 ms, acquisition matrix = $256 \times 256 \times 176$, flip angle = 7°; 1 mm $\times$ 1 mm $\times$ 1 mm voxel size). Functional images were collected using a T2*-weighted echo planar imaging (EPI) sequence sensitive to blood oxygen level dependent (BOLD) contrast (TR = 2000 ms, TE = 30 ms, image matrix = 64×64, FOV = 216 mm, flip angle = 80°, voxel size 3 mm $\times$ 3 mm $\times$ 3 mm, 36 axial slices).

Resting State fMRI Analysis
The first 10 volumes were discarded to allow the magnetisation to approach a dynamic equilibrium and for the participants to get used to the scanner noise. Part of the data pre-processing, including slice timing, head motion correction (a least squares

approach and a six-parameter spatial transformation) and spatial normalization to the Montreal Neurological Institute (MNI) template (resampling voxel size of 3 mm $\times$ 3 mm $\times$ 3 mm), were conducted using the SPM5 and Data Processing Assistant for Resting-State fMRI (DPARSF, Chao-Gan and Yu-Feng, 2010). A spatial filter of 4 mm FWHM (full-width at half maximum) was used. Participants showing head motion above 3.0 mm of maximal translation (in any direction of *x*, *y*, or *z*) or 1.0° of maximal rotation throughout the course of scanning would have been excluded; this was not necessary.

After pre-processing, linear trends were removed. Then the fMRI data were temporally band-pass filtered (0.01–0.08 Hz) to reduce the very low-frequency drift and high-frequency respiratory and cardiac noise (Biswal et al., 1995). Amplitude of low frequency fluctuations (ALFFs) analysis (Yang et al., 2007; Zang et al., 2007) was performed using DPARSF (Chao-Gan and Yu-Feng, 2010). We chose ALFF analysis since it is a commonly used metric with high test-retest reliability (Zuo and Xing, 2014). The time series for each voxel was transformed to the frequency domain using fast Fourier transform (FFT), and the power spectrum was obtained. Because the power of a given frequency is proportional to the square of the amplitude of this frequency component in the original time series in the time domain, the power spectrum obtained by FFT was square rooted and then averaged across 0.01–0.08 Hz at each voxel. This averaged square root is the ALFF (Zang et al., 2007). The ALFF of each voxel was divided by the individual global mean of ALFF within a brain-mask, which was obtained by removing the tissues outside the brain using software MRIcro (by Chris Rorden[2]). A height threshold of $p < 0.001$ was applied to the *t* maps.

We performed the same ALFF analysis on the fMRI data acquired during the extended-duration resting-state sessions except that we removed the images between the onset of the DES beep and 2 min thereafter to exclude activity elicited by tone presentation and the subsequent motor activation while the subjects were jotting down notes. However, we also conducted the ALFF analysis on all extended-duration resting state data, including the 2 min after each beep.

Results

Characterization of Participants
Table 3 shows the percentages of each of the five frequent phenomena (5FP) that had been described by Heavey and Hurlburt (2008). These percentages are shown for each participant, divided into pairs of columns for the natural environment sampling and the in-scanner resting state sampling. We begin by considering the natural environment.

Characteristics of Our Participants in their Natural Environments
The study design called for each participant to undergo 4 days [or more if the interviews called for it; one participant (#3)

---

[1]www.adni-info.org

[2]https://www.nitrc.org/projects/mricron

TABLE 3 | Natural environment and in-scanner resting-state inner experience characteristics (5FP[a] percentages[b]) for each participant.

Participant	Inner speaking		Inner seeing		Unsymbolized thinking		Sensory awareness		Feelings	
	Nat. env.	Scanner rest. st.	Nat. env.	Scanner rest. st.	Nat. env.	Scanner rest. st.	Nat. env.	Scanner rest. st.	Nat. env.	Scanner rest. st.
1 ($n = 13$[c])	8	17	85	19	0	8	85	75	23	3
2 ($n = 16$)	13	22	0	22	6	3	94	75	13	3
3 ($n = 22$)	23	53	0	25	5	17	68	47	14	3
4 ($n = 18$)	39	39	6	42	11	33	6	19	72	22
5 ($n = 12$)	8	14	33	67	25	39	75	78	25	11
Mean	18.2	29.0	24.8	35.0	9.4	20.0	65.6	58.8	29.4	8.4
χ^2 ($df = 4$)	6.99	19.09	47.54	24.57	6.02	22.01	35.56	38.18	21.00	13.82
p (for χ^2)	0.13	0.001	0.0001	0.0001	0.20	0.001	0.000001	0.000001	0.001	0.01
r ($df = 3$)	0.73[d]		−0.09		0.84		0.93		0.95	

[a]5FP = "five frequent phenomena" as discussed by Heavey and Hurlburt (2008). [b]Percentages do not add to 100 because experiences can have more than one characteristic. [c]Number of natural environment samples. Each participant had 36 in-scanner samples. [d]Pearson r between Nat. env. and Scanner rest. st. across participants.

had five natural environment days] of DES sampling in their natural environments, for a potential maximum of 18 samples (24 for participant #3) after discarding the first day. The first column of **Table 3** shows (in parenthesis) the number of natural environment samples we actually obtained for each participant after discarding the first day. The number of natural environment samples varied from participant to participant because training is individualized.

The remaining columns of **Table 3** show the percentages of each of the 5FP for each participant. For example, in the natural environment, participant #1 had 1 instance of inner speaking out of his 13 natural environment samples, so the upper left cell shows $100(1/13) = 8\%$.

The **Table 3** row labeled "Mean" shows that, on average, sensory awareness was our participants' most frequently occurring phenomenon in the natural environment (65.6; we consider in-scanner percentages below). Our participants thus had more frequent sensory awareness than might be expected from Heavey and Hurlburt's (2008) stratified natural environment sample, where only 22% of samples involved sensory awareness. Inner speaking (18.2%), inner seeing (24.8%), and feelings (29.4%) each occurred in the natural environment at about the same frequency found by Heavey and Hurlburt (26, 34, and 26%, respectively). Unsymbolized thinking occurred at a somewhat lower frequency in the natural environment than might be expected from Heavey and Hurlburt (9.4% instead of Heavey and Hurlburt's 22%).

Our Participants Differ from each other in their Natural Environments

Now we ask whether our participants differ from each other on the 5FP (Heavey and Hurlburt, 2008, had reported large individual differences in 5FP characteristics). This is an exploratory study, so we conducted a separate examination of each of the 5FP characteristics. We discovered, for example, that our five participants had very different percentages of sensory awareness when sampled in the natural environment ($\chi^2 = 35.56$,

$p = 0.000001$, $df = 4$). Similar analyses for the other four 5FP phenomena revealed substantial individual differences for inner seeing and for feelings; the inner speaking and unsymbolized thinking differences were smaller. Some individual differences among our participants are large: for example, the natural-environment frequency of sensory awareness ranged from 6% (for participant #4) to 94% (for participant #2); the frequency of inner seeing ranged from 0% (for participants #2 and #3) to 85% (for participant #1), and so on. We performed subsequent analyses in a variety of ways, always with the same results. For example, participant 1 is left handed; if we exclude him from the data, we similarly conclude that the remaining four right-handed people are different from each other in the same ways as all five are different.

Our Participants Differ from each other in the In-Scanner Resting State Sessions

We now consider the in-scanner resting state percentages shown in **Table 3**, asking whether our participants differed from each other in the scanner resting state. We performed the same chi-squared analyses as just discussed, finding that participants in the scanner differed greatly from one another on the frequencies of all five of the 5FP phenomena. We conclude that our participants' inner experiences are different from each other in the resting state. These differences are large: for example, the resting state frequency of sensory awareness ranged from 19% (for participant #4) to 78% (for participant #5); the frequency of inner seeing ranged from 19% (for participant #1) to 67% (for participant #5), and so on. We again performed subsequent analyses in a variety of ways, always with the same results. These results were similar to and more striking than the natural environment differences (probably because the natural environments themselves differed greatly from one participant to the next, and because the iterative nature of DES makes early sampling days (which, in this study, were all natural environment samplings) results relatively unreliable compared to later sampling days (which were all in the scanner).

Our Participants' Experiences in the Resting State Sessions are Broadly Similar to their Experiences in the Natural Environment

We now turn to consider whether a participant's in-scanner resting state percentages shown in **Table 3** are similar to that participant's percentages in their natural environments. We display an exploratory/descriptive Pearson correlation for each of the 5FP across participants ($df = 3$) in the last row of **Table 3**; these correlations are all very high except for inner seeing, which is approximately zero. The high correlations between in-scanner resting state and natural environment does not imply that experience frequency is the same in the scanner and the natural environment. It implies that participants who had a relatively high frequency of a particular characteristic in the natural environment also have a relatively high frequency in the scanner. For example, there were, overall, fewer feelings in the scanner (8.4%) than in the natural environment (29.4%), but the participants who had the most feelings in the natural environment (#4 and 5) also had the most feelings in the scanner. The same, but opposite direction, is true for unsymbolized thinking: there was more unsymbolized thinking in the scanner (20.0%) than in the natural environment (9.4%), but the participants who had the most unsymbolized thinking in the natural environment (#4 and 5) also had the most unsymbolized thinking in the scanner.

The distributions of experiential percentages are sometimes very skewed, which may inflate a Pearson correlation, so we conducted the same analyses using Spearman rank correlations, which are not affected by skew. The Spearman correlations (0.87, −0.05, 0.70, 0.67, and 0.89, respectively) are very similar to the Pearson correlations shown in the bottom row of **Table 3**, suggesting that the correlations between natural environment and resting state phenomena were not due to the extremity of the values.

Thus our data suggest that for these participants, inner experience in the resting state is quite strongly related to inner experience in the natural environment except for the case of inner seeing. We re-emphasize the exploratory nature of these correlations given the small sample size.

Are Extended-Duration Resting State Sessions Experientially Similar to Typical Resting State Sessions?

Next we ask whether our extended-duration resting state sessions are, broadly speaking, similar to typical (5- or 10-min) resting-state sessions. Our resting state sessions were extended in two ways: there were multiple sessions (nine) rather than the more typical one, and each session had a long duration (25 min) rather than the more typical 5 or 10 min; therefore we must explore the potential impact of each of those two kinds of extended duration.

There was Little Drift Across the Multiple Resting State Sessions

First, we investigate whether the 5FP experiential frequencies altered or drifted across this study's multiple (nine) resting state sessions. **Table 4** shows the percentage of samples in each 5FP category, aggregated across participants, displayed by session number. There were four samples for each of the five participants in each session, thus $4 \times 5 = 20$ samples per session. Across all participants in the first session, there were, for example, four instances of inner speaking, so $4/20 = 20\%$ is displayed in the upper left (session 1/inner speaking) cell of **Table 4**. As shown in the middle panel of **Table 4**, a chi-squared test of proportions suggests that our participants' frequency of inner speaking was independent of session ($\chi^2 = 7.68$, $p = 0.47$, $df = 8$). The results are similar for the remaining four 5FP characteristics, as shown in the "All sessions $\chi^2 (df = 8)$" rows of **Table 4**. These χ^2 statistics are all very small (and the corresponding p values large), far from indicating any systematic differences, except for feelings, where $\chi^2 = 20.95$ ($p = 0.01$).

We also asked whether there was a consistent trend across sessions by computing an exploratory/descriptive Pearson correlation between the percentage and the session number; the "r ($df = 7$)" row of **Table 4** shows that those correlations are quite small. The strongest correlation (0.56) was for inner seeing, indicating that inner seeing increased in frequency across scanner sessions.

The first of the nine resting state sessions is arguably the session most similar to a typical one-shot resting state study. As shown in the bottom panel of **Table 4**, we asked whether the 5FP characteristics in the first 25-min resting state session differed from the 5FP characteristics in the remaining eight 25-min resting state sessions. There were no large differences, although there was somewhat more unsymbolized thinking in the first session.

TABLE 4 | Inner experience characteristics (5FP percentages[a]) aggregated across participants by scanner resting-state session.

Session	Inner speaking	Inner seeing	Unsymbolized thinking	Sensory awareness	Feelings
1	20	25	40	50	5
2	30	20	15	40	5
3	45	40	15	70	0
4	40	45	5	75	15
5	15	20	25	70	0
6	25	40	20	65	15
7	35	45	20	60	0
8	30	35	25	60	30
9	20	45	15	40	5
Mean	28.89	35.00	20.00	58.89	8.33
All sessions					
χ^2 ($df = 8$)	7.68	7.91	9.38	11.06	20.95
p (for χ^2)	0.47	0.44	0.31	0.20	0.01
r ($df = 7$)	−0.16	0.56	−0.21	−0.03	0.34
1st session					
χ^2 ($df = 1$)	0.87	0.99	5.63	0.73	0.33
p (for χ^2)	0.35	0.32	0.02	0.39	0.57

[a]*Percentages do not add to 100 because experiences can have more than one characteristic.*

We performed these χ^2 analyses in several different ways, all discovering approximately the same level of independence. For example, we also asked whether the last session differed from the first eight; and whether the first four sessions differed from the last four. None of these tests (not reported here) showed large differences.

Taken together, these results indicate that there were not large or systematic shifts in experience across the nine resting state scanner sessions.

There was Little Drift within the Long Duration Resting State Sessions

Second, we investigate whether our participants' 5FP experiential frequencies altered or drifted within each of this study's long duration (25-min) resting state sessions. **Table 5** shows the aggregation of samples from the first half of each session and those from the second half. For example, we asked whether those aggregates differed for inner speaking and found that there was somewhat more inner speaking in the first half of sessions (37% vs. 21%; $\chi^2 = 5.3$, $p = 0.02$, $df = 1$). The other 5FP characteristics had no large differences, as shown in the bottom two rows of **Table 5**.

Furthermore, we performed all the analyses described for **Tables 2** and **3** within each participant singly, finding similar levels of independence within participants. The p-values for individual participants when performing the kind of analysis shown in **Table 4** are 0.71, 0.56, 0.47, 0.21, and 0.83. Similarly, the individual participant p values comparable to the **Table 5** analysis were 0.68, 0.54, 0.77, 0.02, and 0.98. There is one small p value in those studies: participant #4 can be characterized as having less inner speaking and more inner seeing aggregated across the second halves of all sessions. That is, however, the only relatively large frequency difference.

Thus our data suggest that there are not large or systematic shifts in experience within the nine resting state scanner sessions.

Are Extended-Duration Resting State Sessions Neurophysiologically Similar to Typical Resting State Sessions?

We have explored whether there were *experiential* characteristics that would lead us to suspect that our extended-duration resting state sessions were substantially different from typical resting state sessions, and concluded that there were not. Now we ask whether there were *neurophysiological* characteristics that would

suggest a difference between extended-duration resting state sessions and typical resting state sessions. The small sample size makes customary statistical analysis of fMRI data unsatisfactory, but we can examine in broad strokes whether the brain activity we recorded is similar to the default mode brain activity generally understood to be ongoing in typical resting state scanner studies. As illustrated above in **Figure 1**, our design had a standard 5-min resting state scan in the *introduction/pre-DES resting state* phase and another standard 5-min resting state scan in the *final/post-DES* phase. Our resting state sessions in the *in-scanner DES sampling* phase used typical resting state instructions except for the times participants spent jotting down notes in response to the DES random beeps. As shown in **Figure 2**, we computed ALFF from the data of each participant for each of the three phases, arbitrarily excluding 2 min beginning with the onset of each beep in the *in-scanner DES sampling* phase to exclude brain activity related to the beep and subsequent motor activity. In all 15 plots we find higher ALFF values in the regions typically activated in resting state studies, namely in superior and inferior anterior medial frontal regions, lower precuneus, and posterior lateral parietal cortex. On the possibility that excising the 2 min following each beep creates an artifact, we computed the same 15 plots including the 2 min after each beep in the analysis; those plots are nearly identical to those displayed in **Figure 2** (and are available from the authors).

Phenomena in the Resting State

We inquire now about the phenomena that DES apprehended during the extended-duration resting state scanner sessions. We begin by recalling one of the main findings of **Table 3** above: there are great individual differences in people's resting state experience. The resting state frequency of inner speaking ranged from 17 to 53%; inner seeing ranged from 19 to 67%; sensory awareness ranged from 19 to 78%. Thus our data suggest that there is *no* single kind of resting state phenomenology.

To give a glimpse into the kinds of experience that are found in the resting state, we provide a few typical examples, one from each participant. The entire set of experiences is available from the authors.

1. (Jack, participant 1, sample 8.3) Jack is rubbing with his thumb the fabric that holds the writing board, and he feels the snagging of the fabric on his thumb. This is more a sense in his thumb than of the fabric. He is involved with determining whether this is wool or synthetic, but that is a part of the sensation, not a cognitive act. (Sensory awareness).
2. (Lara, participant 2, sample 7.7) Lara is looking at the edges of the scanner mirror—left bottom corner, and sees two of them, layering. Simultaneously she is hearing herself say, to no one in particular, "I really want to talk to you." The voice is recognized to be her own, expressed in her own natural way; however, the vocal characteristics are not of her own voice but of some female voice that she doesn't recognize. The wrongness of the vocal characteristics was noted only retrospectively—at the moment of the beep, experientially, Lara simply hears herself talking. She knows who the "you" is in this sentence, but the sentence is not directed to that person.

TABLE 5 | Inner experience characteristics (5FP percentages[a]) in first half or second half of scanner sessions.

Session part	Inner speaking	Inner seeing	Unsymbolized thinking	Sensory awareness	Feelings
First half	37	31	18	60	6
Second half	21	39	22	58	11
χ^2 ($df = 1$)	5.30	1.20	0.56	0.09	1.82
p	0.02	0.27	0.46	0.76	0.18

[a]*Percentages do not add to 100 because experiences can have more than one characteristic.*
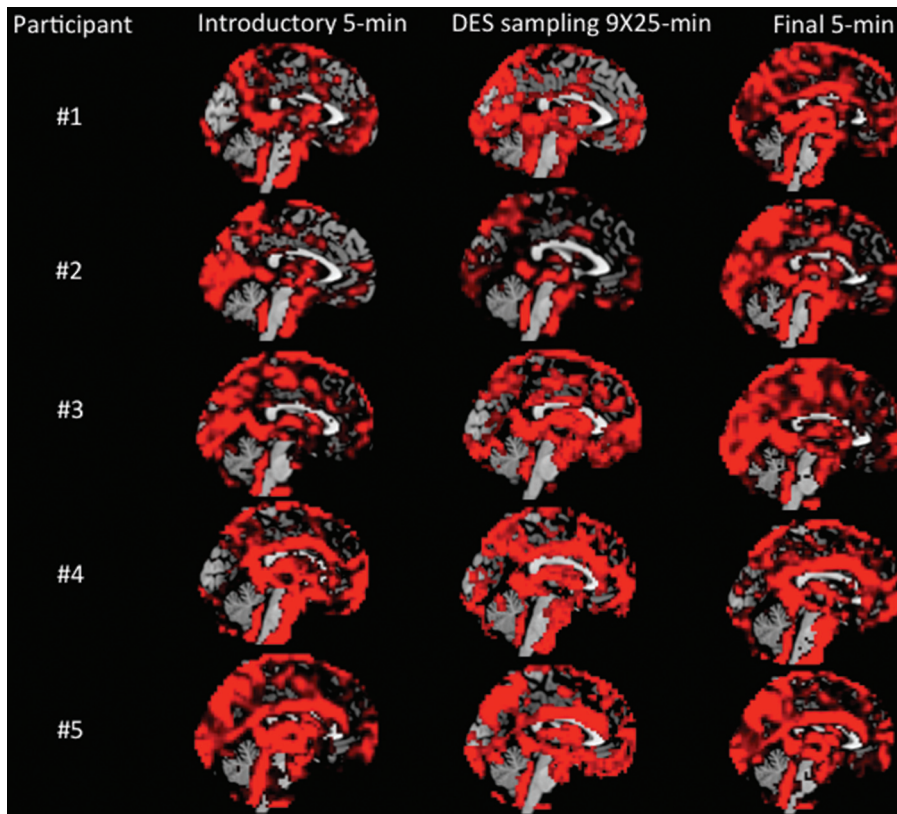
FIGURE 2 | Amplitude of low frequency fluctuation (ALFF) maps for each participant in the three conditions.

She is also seeing her hands. (Sensory awareness and inner hearing with idiosyncratic characteristics).

3. (Otto, participant 3, sample 5.5) Otto is saying in inner speech "I just *turned* 30." He doesn't know why he is saying that or to whom, but he is clear that that is what he is experiencing and that he is emphasizing the word "turned." (Inner speaking).

4. (Susan, participant 4, sample 6.6) Susan innerly sees the actress Sigourney Weaver in a cryogenic tank from the movie *Alien*. She sees Sigourney's face from above, below the glass window of the tank—the rest of Sigourney's body is vaguely or blurrily present. Mostly Susan is searching for the word used in the movie: cryogenic chamber, cryogenic tank, etc., waiting for the right word to appear. This is primarily a state of suspended animation, waiting for the word—she does not see or hear pieces of words, etc. (Inner seeing and an uncategorized experience of cognition).

5. (Jane, participant 5, sample 5.8) Jane is aware of the fidgetiness of her upper trunk, arms, and eyes; they want to move. She is also simultaneously looking at the gray plastic piece of the scanner, noticing especially the dark grayness of it. (Two separate, simultaneous sensory awarenesses).

It can be seen from these examples (and from the entire data set) that resting state experience is active and involved. The resting state is *not* a state of phenomenological rest or suspension, but is typically engaged and explicit. It is sometimes simple but is

often complex and multilayered. Sensory awareness can involve any of the senses, can be actual or imaginary, and can be tied to the environment or distant from it. Inner speaking can be self-directed, other directed, or neither; can be in one's own voice or someone else's; can be meaningful or relatively meaningless. Similar characterizations of the diversity of experience can be made within inner seeing and feelings.

Comparing Questionnaire and DES Sampling

We have seen that there were large individual differences in DES-sampled inner experience but small differences in inner experience across the resting state sessions and within sessions. We now inquire about the adequacy of retrospective accounts of resting state experience. Recall that this study began with a typical 5-min resting state study, immediately followed by the participant's filling out the ReSQ (Delamillieure et al., 2010), a questionnaire designed to characterize inner experiences in the scanner. We now compare the ReSQ (retrospective questionnaire) results with the 5FP (as sampled by DES) results.

The ReSQ is not aimed directly at the 5FP, but two of the ReSQ items seem directly comparable and one is indirectly comparable. The ReSQ presents visual-analog scales which ask participants to rate the resting-state frequency of "visual mental imagery," which can be taken to be comparable to 5FP inner seeing, and "inner language," which can be taken to be comparable to 5FP inner speaking. Additionally, the ReSQ asks participants to rate

"somatosensory awareness," which can be taken to be comparable to a subset of the 5FP sensory awareness. Sensory awareness can be visual, auditory, and so on, as well as bodily, so we split the 5FP sensory awareness into bodily sensory awareness and other sensory awareness, and compared the ReSQ somatosensory awareness to bodily sensory awareness. Then we could compare the three ReSQ characterizations to the sampling frequencies. The ReSQ does not inquire about anything related to the two remaining 5FP (unsymbolized thinking or feelings).

Table 6 presents the ReSQ percentages side by side with the DES (sampling) percentages for the three comparable characteristics for each participant. For example, participant #1 used the visual analog scales of the ReSQ to report that he had spent 40% of his 5-min resting-state time engaged in inner language. His DES sampling showed that 17% of his 36 DES in-scanner samples involved inner speaking.

The "Max discrepancy" row of **Table 6** shows the largest difference between a participant's ReSQ percentage and DES percentage. It can be seen that some of these discrepancies are very large; for example, the maximum discrepancy for inner speaking (Participant 5's) was 81% (her estimate on the ReSQ was 95% minus her 14% obtained by DES).

After participating in 13 DES sampling periods and their 13 expositional interviews, each participant again underwent (in the *Final/post-DES* phase) a standard 5-min resting state scan, immediately followed by another administration of the ReSQ. The second ReSQ maximum discrepancies were substantially smaller than the original ReSQ maximum discrepancies (29.5, 20.5, and 23.5, respectively).

High Fidelity in a Single Case: Jane's Inner Experience

We have established that, at least for this small sample, people's retrospective or general characterizations of their resting state experience are not to be accepted as faithful accounts of actual experience while in the scanner. We have not, however, described that actual experience. We now provide a glimpse into the experience of one participant, described in as high fidelity as we can muster.

We choose to describe participant #5, Jane, because she exemplifies those (perhaps a majority) who are substantially mistaken about their own inner experience. Jane believed, prior

to participation, that she talked to herself nearly all the time (as reflected by her self report, by a 90% inner speech rating on a questionnaire not reported here, and by her ReSQ inner speech rating of 95%); however, sampling revealed that she talks to herself only rarely. In the natural environment she had only one example of inner speaking:

(Jane sample 3.4) Jane was in the U-Bahn, sitting on a bench waiting for a train. She is hearing the sound of kids talking—lots of them; she hears their chatter—loud and clear—how they sound, not what they are saying. She is also writing "I cannot see" and saying those words in inner speech, but more attending to the writing.

The inner speaking here is only the third most prominent feature of her experience (after hearing the chatter and attending to the writing).

In the scanner, there was one straightforward example of Jane's inner speaking:

(Jane sample 5.5) Jane is innerly asking her friend Sharon where she lives in Berlin. Jane is asking this in Swedish (which is both her native tongue and Sharon's) at a pace that she described as similar to ordinary speaking in that language. Simultaneously Jane, whose eyes are closed, is seeing ambiguous dark shapes in black, yellow, and gold. The two experiences (speaking and seeing colors) are approximately equal in significance/attention.

We counted two samples as containing inner speaking where the speech might not be considered "I thought in words" by most people:

(Jane sample 8.2) Jane is looking at the writing tablet in the mirror and mainly attending to its dark, rectangular shape, which takes up most of the mirror. She can see the piece of paper on it, which is small. The white paper stands out against the dark background. At the moment of the beep she is also innerly saying "Oh yeah!" and thinking that she should email the kindergarten that she had been intending to volunteer at. The thought about the kindergarten is not worded or accompanied by any pictures.

"Oh yeah" was clearly innerly spoken, so we counted it as inner speaking, but the thought was about kindergarten volunteering, and "Oh yeah" was more an accompanying ejaculation (possibly expressing a motivational function of inner speech; McCarthy-Jones and Fernyhough, 2011).

(Jane sample 7.6) Two things are ongoing simultaneously: (1) Jane is seeing the half-moon shape at the top of the scanner mirror. She sees everything that is out there – the people, the tops of computers, etc., particularly noticing the blue/black striped shirt of the person who is walking back and forth. Everything is clear; the striped shirt is more central and more detailed. (2) Jane is imagining herself reading a story aloud to kids. She is growling like a wolf or a bear, and she feels her eyebrows rising for effect, and she feels her fingers straightening out and expanding for terrific effect (in reality, there is no bodily movement). There is a sense of the kids present, but not particular kids (probably kids that she knows).

The growling was oral, so we (liberally) counted it as inner speaking.

TABLE 6 | Resting-State Questionnaire (first administration) percentages compared to DES sampling percentages.

Participant	Inner speaking		Inner seeing		Bodily sensory awareness	
	ReSQ	DES	ReSQ	DES	ReSQ	DES
1	40	17	14	19	38.5	39
2	40.5	22	57.5	22	31.5	36
3	69	53	11	25	78.5	33
4	20	39	60	42	20	8
5	95	14	85	67	42	11
Max discrepancy	81.0		35.5		45.5	

There was one sample where inner speaking may have been ongoing:

(Jane sample 6.6) Jane is thinking that her hair was a mess and is simultaneously innerly seeing herself in the mirror in her bathroom at home, a remembering of what she had seen earlier. The thinking portion: she is thinking that her hair is a mess (even though she doesn't see her hair very completely in the mirror); the thinking involves some hints of words that might be like "Oh man my hair is a mess." There were not fully formed words, although neither were words completely absent. The seeing portion: the seeing of herself in the mirror includes the seeing of the bathroom, seen through her own eyes, including the mirror, the bright light, the sink, the toilet, and part of her body. There is no specific focus to the image. The thought (of hair being a mess) was the main focus, about 60/40.

There were two samples which involved Jane's voice being innerly heard (rather than spoken) as an aspect of a recollection. Here is one:

(Jane sample 5.6) Jane is imaginarily replaying a conversation she had with her friend Sharon. Jane innerly sees the floor, sees Sharon in her inner peripheral vision, and hears herself say "If you look at me I don't look Swedish." Mostly Jane is paying attention to her own words, which she hears (note that in the original conversation she had spoken the words, but now she hears them, not speaks them).

There were two samples where words were present but which had no experienced meaning. Here's one:

(Jane sample 7.1) Jane innerly hears herself saying "thinks about what used to be there," said in her own voice, as clear as hearing someone else externally speaking. She can't control the speech and it doesn't present itself as meaningful speech (that is, she doesn't know *who* thinks, where *there* is, or what *used to be* there). As an equal part of her experience (50:50) Jane simultaneously innerly sees a long piece of something pale and white; simultaneously she somehow knows it to be a log, even though it doesn't look like a log. That is, she does not see a log. The words she can hear are understood to be related to the image, but she doesn't know why/how.

If one counts all those samples which contain the experience of words or hints of words—frank inner speech, inchoate inner speech, inner hearing, and so on—the frequency of Jane's words is 14% in the natural environment and 25% in the scanner, far smaller than her prior-to-sampling self-understanding (90 or 95%). Part of the discrepancy might be accounted for by the fact that Jane frequently engaged in unsymbolized thinking (Hurlburt and Akhter (2008a)): the direct, unambiguous experience of thinking that is not accompanied by words, images, or other symbols (25% in the natural environment and 39% in the scanner). Here, are two examples from the scanner:

(Jane sample 5.1) Jane is wondering whether it is possible to 100% do two things at once (a reference to something RTH had said a few days before). There are no words or symbols, even though she is quite specific about the "100%" part and the "at once" part, etc. She is simultaneously also slightly attending to the scanner sound.

(Jane sample 8.7) Jane's eyes have closed in a blink and she sees a negative image of the half-moon scanner scene: what is actually white she sees dark, and what is actually dark she sees light pale yellow. This seeing starts where the actual scene exists and then floats downward. Simultaneously she is thinking a little bit about Master's programs: how courses are selected, how do you decide what thesis to write, how do you write it, etc. This is all without words or images and are all aspects of one thinking.

Hurlburt and Akhter (2008a) reported that many people who have frequent unsymbolized thinking initially believe that such thinking is impossible and (mistakenly) understand themselves to be thinking in words.

Furthermore, Jane's most frequent kind of experience was sensory awareness (75% in the natural environment and 78% in the scanner). We have given several examples above (samples 3.4, 5.5, 5.8, 7.1, 7.6, and 8.2). Here is another (which includes three simultaneous sensory awarenesses):

(Jane sample 8.8) Jane is seeing the knuckles on her left hand, especially the silvery blue shades of her knuckles caused by the blue light. Simultaneously she hears the incredibly loud noise of the scanner, primarily in her right ear, and she feels her upper body vibrate in sync with that noise.

Hurlburt et al. (2009) reported that many people who have frequent sensory awareness initially (mistakenly) believe that such experience is explicitly cognitive, and (mistakenly) understand themselves to be thinking in words. For example, it is likely that, before sampling, had Jane had the experience described in sample 8.8, she would have (mistakenly) believed that she was saying to herself that the silvery blue shade of her knuckles was interesting.

Over the course of sampling, Jane came to the realization that she frequently had sensory experiences without cognitive overlay and that she frequently had thoughts that did not involve words (what we call unsymbolized thinking). Perhaps as a result, when she characterized her frequency of verbal thinking on the second (end of sampling) administration of the ReSQ, she estimated her resting-state inner speech percentage to be 22.5%, much lower than her first ReSQ estimate of 95%.

Discussion

We have noted the importance of exploring the experiential phenomena of the resting state, an undertaking that has not been accomplished heretofore because of methodological challenges. We address those challenges using DES, a technique designed to explore experience in high fidelity, in an extended duration resting state fMRI study that involved multiple (nine) sessions, each of long duration (25 min), for a total of 225 min in the scanner rather than the more usual 5 or 10 min.

Because of the intensity of the effort (for each participant, 11 scanner sessions and 13 DES expositional interviews, each with two or more highly skilled interviewers), we used in this exploratory study only five participants. This study is therefore best considered a small first step in an important direction rather than a definitive investigation. All the results that we have

reported above and will discuss below must be understood in the context of weighing the risks of small-n descriptive studies against their benefits (Hurlburt, 2011a, Chap. 21). The results suggest that the neurophysiological sophistication of the scanner can be profitably combined with the phenomenological sophistication of DES, and that opens substantial possibilities for issues that are central to neuroscience and psychology in general.

Our results indicate that there were not large or systematic shifts in experience across the multiple (nine) sessions or within the long-duration (25 min) sessions. Furthermore, our brain activation results showed that the ALFF patterns in each participant in each of the three phases (the initial 5-min resting state session, the concatenated extended-duration resting state sessions, and the final 5-min resting state session) all showed activation in superior and inferior anterior medial frontal regions, lower precuneus and posterior lateral parietal cortex as would be expected from typical resting state studies. Taken together, these results suggest that our extended-duration resting state sessions were experientially and neurophysiologically broadly similar to typical resting-state sessions.

We found substantial individual differences in resting-state experience across our participants. For example, the resting-state frequency of sensory awareness ranged from 19 to 78%; inner seeing ranged from 19 to 67%; inner speaking ranged from 14 to 53%; and so on. We found similar wide ranges in the natural environment, as did Heavey and Hurlburt (2008). It therefore seems that our five participants were, in broad strokes, similar to what might be expected in a sample of volunteers for psychological or neuropsychological studies. We note that the import of this finding is not diminished by the small sample size: we have documented substantial experiential differences, calling into question the frequently held assumption of universal experiential characteristics.

With the exception of inner seeing, we found a substantial relationship between a person's experiential frequencies in the natural environment and that person's in-scanner resting state. At least as measured by the 5FP, our participants apparently engaged in approximately the same forms of experience in the scanner as they do in their own everyday environments—those who have frequent sensory awareness in the natural environment also have frequent sensory awareness in the scanner, and so on. This finding suggests that the term "resting state" may have two unfortunate connotations: that people are psychologically at rest and that there is one state in which they find themselves. Our data suggest that the default mode network (DMN) may be activated because people are *engaging in their usual kinds of spontaneous, everyday experience* in the scanner, the same kind of experience they would engage in if they were actively participating in their own wide-ranging everyday undertakings. That is, the DMN may be active when people experientially do what they usually do (whether resting or not), and is suppressed when the person is instructed by an experimenter to do something foreign or unnatural to that individual. Perhaps scientists characterizing their participants when not engaged in tasks should refer to "unconstrained activity" rather than to the "resting state."

We emphasize the desirability of replication by others with larger sample sizes. The present study is unique in that it gathers data both in the natural environment and in the scanner, thus affording the first opportunity to compare natural-environment to in-scanner experience. We note the important caveat that the natural environment data gathered here are experiential, so it requires an extrapolation to infer that brain activation, like experience, is similar in the natural environment and the scanner.

As we saw in **Table 1**, researchers have provided a wide variety of characterizations about phenomena in the resting state. Some have noted the verbal nature of this experience (see **Table 1** rows 1 and 2). The present study provided $5 \times 36 = 180$ glimpses of experience in the resting state; of those, approximately 58 (32%) involved words of any kind, whether innerly spoken, innerly heard or imagined in any other way. The determination of whether a particular sample involves words is not perfectly reliable, so some might say our 32% over-represents or under-represents the frequency of words to some degree, but it is safe to conclude that most of our participant's sampled resting-state experiences were *not* verbal.

Some researchers have characterized resting state experience as involving planning (**Table 1** rows 3 and 4). Of the present study's 180 glimpses, approximately 39 (22%) could be said to involve planning when "planning" was defined in a very inclusive way. Here again, the determination of whether planning was part of a particular sample is far from perfectly reliable, but it is safe to conclude that most of our sampled experiences did *not* involve planning for the future.

Some researchers have characterized resting state experience as involving enhanced attention (**Table 1** rows 5 and 6). Our participants did indeed have frequent specific awareness of inner or external stimuli. However, their sensory awarenesses were not *enhanced* or *increased*: our participants had very frequent specific and engrossing sensory awareness in their own natural environments—in fact, the frequency of such experience was slightly *higher* in the natural environment (65.6%) than in the scanner (58.8%). Such naturally occurring sensory awarenesses are very frequently overlooked by people who engage in such phenomena frequently (Hurlburt et al., 2009).

Some researchers have characterized resting state experience as involving an internal monitoring (**Table 1** rows 7, 8, and 9). However, others have held that it involves an external monitoring (**Table 1** row 5), whereas yet others have held that it is an alternation of internal with external (**Table 1** row 6). Our results suggest that actual experience is sometimes inwardly directed, sometimes externally directed, and sometimes neither, with no direction overwhelmingly predominant.

Some researchers have characterized resting state experience in ways that we find difficult to parse in experiential terms (**Table 1** rows 10 and 11). We invite them to peruse our glimpses (all available from us) and draw appropriate conclusions.

We now ask whether the questionnaires designed to investigate resting-state phenomena can do so in high fidelity. The ReSQ asks participants to characterize their experience into five categories (visual imagery, in inner language, in

somatosensory awareness, in inner musical experience, and in the mental manipulation of numbers) using instructions that require "that the total score for the five types of activity had to equal 100%" (Delamillieure et al., 2010, p. 566). That instruction presupposes that categories are mutually exclusive: that an experience is either, for example, inner language or somatosensory awareness but not both. Our own participants provide 24 examples (13%) where that mutual-exclusive assumption is far from correct—our Lara (participant 2, sample 7.7) above is one such example (as is example 5, in one respect). Here is another: Otto (sample 7.5) is reciting a poem to himself (Goethe's "Erlkönig") and he is currently on the third line (of four) of the seventh stanza (of eight). He is innerly speaking the line "Mein Vater, mein Vater, jetzt faster mich an" in a soft declarative voice. Simultaneously he is counting the stanza on his left fingers and the line within stanza on his right finger (so he is holding his second finger left hand against the desk to indicate seventh stanza and the third finger right hand against the desk to indicate third line). He is more aware of his right hand (apparently because it is about to advance to the fourth line). Simultaneously he is actively trying not to hear the noise of the scanner. That is, he is *not* merely automatically screening out the noise (which he does successfully on other occasions). He hears the noise but he is actively trying not to attend to it. We conclude that the ReSQ presupposition that characteristics must add to 100% is importantly misguided.

The two other questionnaires aimed at characterizing the resting state (ARSQ, Diaz et al., 2013, and ARSQ 2.0, Diaz et al., 2014; and the NYC-Q; Gorgolewski et al., 2014) also have presuppositions that we believe substantially interfere with their ability to ascertain with fidelity the characteristics of experience. For example, both inquire about thoughts. The ARSQ 2.0 instructions ask participants in an online study to sit quietly in front of their computer screen for 5 min. At the conclusion of the 5 min, a screen appears which reads, "The 5 min of rest are over. Now several statements will follow regarding potential feelings and thoughts you may have experienced during the resting period. Please indicate the extent to which you agree with each statement" (Diaz et al., 2014, p. 2). Statements such as "I thought in words" and "I thought about myself" are to be rated on a "five-point ordinal scale with the labels 'Completely Disagree,' 'Disagree,' 'Neither Agree nor Disagree,' 'Agree,' and 'Completely Agree'" (Diaz et al., 2013, pp. 2–3). The NYC-Q asks participants to report on the content and form of their self-generated thoughts using Likert scales for items that begin "I thought about" *X* or "I thought of" *X* (Gorgolewski et al., 2014, p. 3). However, careful DES interviewing reveals consistently that people use the word "thinking" (or "thought") in highly disparate ways, leading Hurlburt and Heavey to conclude:

> We take it as an axiom that when a DES participant says "I was thinking…," we know nothing whatever about the phenomena of her inner experience. That claim may be highly counterintuitive, because most people, when asked, do in fact define "thinking" as a cognitive event and correctly discriminate between, for example, "thinking" and "feeling" when observing one person solving a math problem and another crying. Our claim is that when people speak of the experience of others, the referent of thinking is some

cognitive process or event, but when they speak of themselves, the referent of thinking is frequently not cognitive and is unspecified and/or unspecifiable.

> The word thinking is arguably the most problematic word in the exploration of pristine experience (Hurlburt and Heavey, 2015, p. 151).

Despite the phenomenological ambiguity of the words "thinking" and "thought," most of the ARSQ items (18 out of 27) and all of the NYC-Q items inquire about thinking as if participants would have a joint understanding about what is being asked. Hurlburt (2011a) held that it requires iterative training to help participants use experiential terminology in consistent ways. For example, suppose that our Jane had had an experience such as her sample 5.8 (fidgetiness noticing the dark grayness of the plastic piece) but had not undergone the DES iterative training. Hurlburt (2011a) would suggest that Jane would very likely have reported herself to have been *thinking* that she wanted to move and *thinking* about the gray plastic piece, rather than to have been engaged in sensory awarenesses of those aspects of her environment.

It might be observed that the DES expositional interviews are themselves retrospective, just like questionnaire reports: both occurred following the participant's exiting the scanner. However, the DES expositional interviews are constrained by the notes jotted down contemporaneously. One might then observe that even those "contemporaneous" notes are actually retrospective—the target experience was a few seconds prior to the note jotting. This criticism holds that retrospection, no matter how short, so disrupts experience that what *seems* like recollection is actually a construction that is unrelated to whatever experience may have been ongoing at the moment of the beep. We think that is unlikely given our careful questioning of hundreds of people in sampling situations, but we accept that it is possible. However, if that possibility is taken seriously, then *all* first-person reports should be excluded from science, because *all* first-person reports require some sort of retrospection. Excluding all first-person reports is an entirely defensible (although we think misguided) scientific strategy. This issue has been discussed at length in Hurlburt and Schwitzgebel (2007) and Hurlburt (2011a).

It might be observed that the natural environment training performed in this study was more inefficient—even accepting that iterative training is necessary—than performing that iterative training in the scanner, where experiences of exactly the sort that are found in the scanner could be examined and clarified. Hurlburt (2011a) holds that training in a target environment is a risky practice. One of the main objects of the iterative procedure is to help participants overcome or bracket their presuppositions about experience. That is always a difficult thing to accomplish, but it is made more difficult the more closely tied the training is to the target experiences. Furthermore, it would undermine the confidence one might have in the results. For example, suppose that in the present study there had been no natural environment training. Further suppose (as actually happened) that the study discovered exceptionally high frequencies of sensory awareness. In that case, the criticism could have been effectively leveled that the high frequency of sensory awareness was an artifact of the

in-scanner training: the noise of the scanner would have been a salient characteristic of the first scanner session, and therefore a large topic of conversation in the first expositional interview. The participant could have gleaned that the interviewer was particularly interested in sensory awareness, and would therefore have been more likely to report sensory awareness in subsequent interviews. However, the study as designed diminishes that criticism. Participants were introduced to sampling in their natural environments, which were no more nor less sensorially salient than usual. The first expositional interviews were not skewed by the experimental situation toward the sensory, and in fact covered a wide range of topics and characteristics absolutely unrelated to scanning features. On the basis of those interviews, we discovered that our participants had sensory interests, but we also conveyed by word and deed to our participants that we were not *particularly* interested in sensory aspects. With that as background, we think that whatever phenomena are discovered in the in-scanner sampling are more believable.

It might be observed, following Ross and Nisbett (1991), that the intensive, multiple interviews such as those we have conducted run the risk of biasing the participant. That is indeed a risk, but Hurlburt has written extensively (Hurlburt and Heavey, 2006; Hurlburt, 2011a) about the many ways DES manages that risk. For one example, the DES procedure is "open-beginninged" (Hurlburt, 2011a): the interviewer does not initially inquire about any phenomenon that has been specified *a priori* (that is, DES does *not* initially inquire about images, or about inner speech, or about any other pre-defined topic) but instead asks for a description of whatever experience, if any, was ongoing at the moment of the beep, and then follows up on the participant's response. Hurlburt (2007, pp. 285–289) has argued that Ross and Nisbett's own analysis shows that "opening up the channel factor" mitigates the obedience risk, and Hurlburt has shown how DES gives multiple and repeated channel-opening instructions:

> For example, we explicitly and repeatedly told (DES participant) Melanie that she could withdraw at any time; that saying "I don't know" or "I don't remember" was a perfectly legitimate response, that we valued her best effort over any predetermined expectation; that it was quite possible that things wouldn't be clear and that that was okay; that the task was perhaps impossible; that we would learn as much or more from her inability to perform a task as we would from her ability to perform it easily; that we much preferred her unexaggerated candor to any attempt to figure out what we wanted to hear; and so on. Not only did we say such things repeatedly, but we meant them sincerely; and not only did we mean them sincerely, I think Melanie recognized that we meant them sincerely. Therefore, by Ross and Nisbett's own argument, we, I think, successfully undermined the channel effect and therefore should *not* expect large obedience effects.
> (Hurlburt, 2007, p. 288)

It might be asked whether the intensive training provided in this study alters the participants so that their resting states are no longer similar to untrained individuals. We do indeed believe that participants become more skilled at apprehending their experience, but for most participants, we think that does not substantially alter the nature of their experience. **Table 4** provides a bit of support for that statement: there is not substantial

experiential drift across the sessions. Replication by multiple methods is desirable.

Even if one accepts the fidelity of the individual observations, it might be observed that because of the small sample size ($n = 5$), generalization to any population is risky. That is indeed true (and we have tried to insert appropriate caveats to that effect throughout), but there are advantages and disadvantages to all approaches in science including large sample size. We take as an example the Diaz et al. (2014) briefly described above, because it is an exemplary study at the state of its art and one which appeared recently in this journal.

Diaz et al. (2014) had a large sample (1444 participants) fill out the ARSQ or ARSQ 2.0 to investigate "mind-wandering experiences," including "verbal thought." The disadvantage of large n is that Diaz et al. (2014) cannot know what any participant intended when endorsing *Agree* to any of the ARSQ statements, for example, to "I thought in words." Participants might endorse *Agree* because they thought in words *frequently*, or because they thought in words *once* but recall it, or because they mistakenly believe themselves to think in words even though they *never* actually thought in words during the experiment. To sort through those important differences requires an iterative method (Hurlburt, 2011a), which in turn requires multiple interviews and multiple skilled interviewers, making large sample sizes impossible.

Questionnaire responses can be validated in a variety of ways (see Alderson-Day and Fernyhough, 2015, for a review of inner speech validation), and such validation is necessary to science. However, Hurlburt (2011a, Chap. 21) argued in favor of a science that values both validity-based and observation-based methods and that recognizes the very different constraints that operate in each realm. Here, for example, we discovered that one of our five participants, Jane, firmly believed herself (prior to participation) to be a nearly-all-the-time inner speaker but was more likely a nearly-none-of-the-time inner speaker. Does that imply that 20% of people are hugely mistaken about their inner speech? Certainly not—such a claim would indeed require a large n study. However, it does show that *some* people (Hurlburt, 2011a, believes that Jane is by no means exceptional) are hugely mistaken. Furthermore, we have shown that a few samples from one person can provide rich insight into the nature of verbal experience—that words are sometimes spoken, sometimes heard, sometimes inchoate, sometimes meaningless. That is a revealed phenomenological richness that cannot possibly be obtained from questionnaires, even from thousands of responses to "Please indicate the extent to which you agree with 'I thought in words'." Such results can be obtained only from idiographic, iterative examination of individuals (Hurlburt, 2011a), and such depth of examination and careful disentangling of language is possible only with small numbers of participants.

This was a small study in terms of number of participants, but a large study in terms of intensity [13 DES expositional interviews, 11 scanner sessions (approximately 275 min) per participant]. The results are provocative: they suggest that high fidelity descriptions of inner experience can be gathered in the scanner; they suggest that there are large individual differences

in inner experience in the scanner; they suggest that experience in the resting state may be characterized as being unconstrained activity. Clearly these results need to be replicated by larger studies and by investigators unrelated to us, but if replicated, they would make significant contributions to scientists' quest to integrate information about experience and brain function in the MRI scanner.

References

Alderson-Day, B., and Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychol. Bull.* 141, 931–965. doi: 10.1037/bul0000021

Andreasen, N. C., O'Leary, D. S., Cizadlo, T., Arndt, S., Rezai, K., Watkins, G. L., et al. (1995). Remembering the past: two facets of episodic memory explored with positron emission tomography. *Am. J. Psychiatry* 52, 1576–1585.

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., and Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.* 104, 322–335. doi: 10.1152/jn.00830.2009

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., and Cox, R. W. (1999). Conceptual processing during the conscious resting state: a functional MRI study. *J. Cogn. Neurosci.* 11, 80–93. doi: 10.1162/089892999563265

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Callard, F., and Margulies, D. S. (2014). What we talk about when we talk about the default mode network. *Front. Hum. Neurosci.* 8:619. doi: 10.3389/fnhum.2014.00619

Callard, F., Smallwood, J., Golchert, J., and Margulies, D. S. (2013). The era of the wandering mind? Twenty-first century research on self-generated mental activity. *Front. Psychol.* 4:891. doi: 10.3389/fpsyg.2013.00891

Chao-Gan, Y., and Yu-Feng, Z. (2010). DPARSF: A MATLAB toolbox for "pipeline". Data analysis of resting-state fMRI. *Front. Systems Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., and Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8719–8724. doi: 10.1073/pnas.0900234106

Delamillieure, P., Doucet, G., Mazoyer, B., Turbelina, M., Delcroix, N., Mellet, E., et al. (2010). The resting state questionnaire: an introspective questionnaire for evaluation of inner experience during the conscious resting state. *Brain Res. Bull.* 81, 565–573. doi: 10.1016/j.brainresbull.2009.11.014

Diaz, B. A., VanDerSluis, S., Benjamins, J. S., Stoffers, D., Hardstone, R., Mansvelder, H. D., et al. (2014). The ARSQ 2.0 reveals age and personality effects on mind-wandering experiences. *Front. Psychol.* 5:271. doi: 10.3389/fpsyg.2014.00271

Diaz, B. A., VanDerSluis, S., Moens, S., Benjamins, J. S., Migliorati, F., Stoffers, D., et al. (2013). The Amsterdam Resting State Questionnaire reveals multiple phenotypes of resting-state cognition. *Front. Hum. Neurosci.* 7:446. doi: 10.3389/fnhum.2013.00446

Fox, K. C. R., and Christoff, K. (2014). "Metacognitive facilitation of spontaneous thought processes: when metacognition helps the wandering mind find its way," in *The Cognitive Neuroscience of Metacognition*, eds S. M. Fleming and C. D. Frith (Berlin: Springer-Verlag), 293–319.

Fransson, P. (2005). Spontaneous low-frequency BOLD signal fluctuations: an fMRI investigation of the resting-state default mode of brain function hypothesis. *Hum. Brain Mapp.* 26, 15–29. doi: 10.1002/hbm.20113

Gilbert, S. J., Dumontheil, I., Simons, J. S., Frith, C. D., and Burgess, P. W. (2007). Comment on "Wandering minds: the default network and stimulus-independent thought." *Science* 317, 43. doi: 10.1126/science.1140801

Gorgolewski, K., Lurie, D., Urchs, S., Kipping, J. A., Craddock, R. C., Milham, M. P., et al. (2014). A correspondence between individual differences in the brain's intrinsic functional architecture and the content and form of self-generated thoughts. *PLoS ONE* 9:e97176. doi: 10.1371/journal.pone.0097176

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100

Gusnard, D. A., and Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat. Neurosci. Rev.* 2, 685–694. doi: 10.1038/35094500

Heavey, C. L., and Hurlburt, R. T. (2008). The phenomena of inner experience. *Conscious. Cogn.* 17, 798–810. doi: 10.1016/j.concog.2007.12.006

Heavey, C. L., Hurlburt, R. T., and Lefforge, N. (2010). Descriptive Experience Sampling: a method for exploring momentary inner experience. *Qual. Res. Psychol.* 7, 345–368. doi: 10.1080/14780880903009274

Heavey, C. L., Hurlburt, R. T., and Lefforge, N. (2012). Toward a phenomenology of feelings. *Emotion* 12, 763–777. doi: 10.1037/a0026905

Hurlburt, R. T. (1990). *Sampling Normal and Schizophrenic Inner Experience.* New York, NY: Plenum Press.

Hurlburt, R. T. (1993). *Sampling Inner Experience in Disturbed Affect.* New York, NY: Plenum Press.

Hurlburt, R. T. (2007). "Russ's reflections," in *Describing Inner Experience? Proponent Meets Skeptic,* eds R. T. Hurlburt and E. Schwitzgebel (Cambridge, MA: MIT Press), 252–291.

Hurlburt, R. T. (2009). Iteratively apprehending pristine experience. *J. Conscious. St.* 16, 156–188.

Hurlburt, R. T. (2011a). *Investigating Pristine Inner Experience: Moments of Truth.* Cambridge: Cambridge University Press.

Hurlburt, R. T. (2011b). Descriptive Experience Sampling, the Explicitation Interview, and pristine experience: in response to Froese. Gould, & Seth. *J. Conscious. Stud.* 18, 65–78.

Hurlburt, R. T., and Akhter, S. A. (2006). The descriptive experience sampling method. *Phenomenal. Cogn. Sci.* 5, 271–301. doi: 10.1007/s11097-006-9024-0

Hurlburt, R. T., and Akhter, S. A. (2008a). Unsymbolized thinking. *Conscious. Cogn.* 17, 1364–1374. doi: 10.1016/j.concog.2008.03.021

Hurlburt, R. T., and Akhter, S. A. (2008b). Unsymbolized thinking is a clearly defined phenomenon: a reply to Persaud. *Conscious. Cogn.* 17, 1376–1377. doi: 10.1016/j.concog.2008.07.004

Hurlburt, R. T., and Heavey, C. L. (2006). *Exploring Inner Experience: The Descriptive Experience Sampling Method.* Amsterdam: John Benjamins.

Hurlburt, R. T., and Heavey, C. L. (2015). Investigating pristine inner experience: implications for experience sampling and questionnaires. *Conscious. Cogn.* 31, 148–159. doi: 10.1016/j.concog.2014.11.002

Hurlburt, R. T., Heavey, C. L., and Bensaheb, A. (2009). Sensory awareness. *J. Conscious. Stud.* 16, 231–251.

Hurlburt, R. T., Heavey, C. L., and Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Conscious. Cogn.* 22, 1477–1494. doi: 10.1016/j.concog.2013.10.003

Hurlburt, R. T., and Schwitzgebel, E. (2007). *Describing Inner Experience? Proponent Meets Skeptic.* Cambridge, MA: MIT Press.

Hurlburt, R. T., and Schwitzgebel, E. (2011). Methodological pluralism, armchair introspection, and DES as the epistemic tribunal. *J. Conscious. Stud.* 18, 253–273.

Kühn, S., Fernyhough, C., Alderson-Day, B., and Hurlburt, R. T. (2014). Inner experience in the scanner: can high fidelity apprehensions of inner

experience be integrated with fMRI? *Front. Psychol. Cogn. Sci.* 5:1393. doi: 10.3389/fpsyg.2014.01393

McCarthy-Jones, S., and Fernyhough, C. (2011). The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Conscious. Cogn.* 20, 1586–1593. doi: 10.1016/j.concog.2011.08.005

Raichle, M. E. (2011). The restless brain. *Brain Connect.* 1, 3–12. doi: 10.1089/brain.2011.0019

Ross, L., and Nisbett, R. E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. Philadelphia, PA: Temple University Press.

Ruby, F. J. M., Smallwood, J., Engen, H., and Singer, T. (2013). How self-generated thought shapes mood—The relation between mind-wandering and mood depends on the socio-temporal content of thoughts. *PLoS ONE* 8:e77554. doi: 10.1371/journal.pone.0077554

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *J. Cogn. Neurosci.* 9, 648–663. doi: 10.1162/jocn.1997.9.5.648

Smallwood, J., Brown, K., Baird, B., and Schooler, J. W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *Brain Res.* 1428, 60–70. doi: 10.1016/j.brainres.2011.03.072

Smallwood, J., and Schooler, J. W. (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Ann. Rev. Psychol.* 66, 487–518. doi: 10.1146/annurev-psych-010814-015331

Wicker, B., Ruby, P., Royet, J.-P., and Fonlupt, P. (2003). A relation between rest and the self in the brain. *Brain Res. Rev.* 43, 224–230. doi: 10.1016/j.brainresrev.2003.08.003

Yang, H., Long, X.-Y., Yang, Y., Yan, H., Zhu, C.-Z., Zhou, X.-P., et al. (2007). Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI. *Neuroimage* 36, 144–152. doi: 10.1016/j.neuroimage.2007.01.054

Zang, Y. F., He, Y., Zhu, C. Z., Cao, Q. J., Sui, M. Q., Liang, M., et al. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain Dev.* 29, 83–91. doi: 10.1016/j.braindev.2006.07.002

Zuo, X.-N., and Xing, X.-X. (2014). Test-restest reliabilities of resting-state fMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009

Inner experience in the scanner: can high fidelity apprehensions of inner experience be integrated with fMRI?

Simone Kühn¹*, Charles Fernyhough², Benjamin Alderson-Day² and Russell T. Hurlburt³

¹ Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany
² Department of Psychology, Durham University, Durham, UK
³ Department of Psychology, University of Nevada, Las Vegas, Las Vegas, NV, USA

To provide full accounts of human experience and behavior, research in cognitive neuroscience must be linked to inner experience, but introspective reports of inner experience have often been found to be unreliable. The present case study aimed at providing proof of principle that introspection using one method, descriptive experience sampling (DES), can be reliably integrated with fMRI. A participant was trained in the DES method, followed by nine sessions of sampling within an MRI scanner. During moments where the DES interview revealed ongoing inner speaking, fMRI data reliably showed activation in classic speech processing areas including left inferior frontal gyrus. Further, the fMRI data validated the participant's DES observations of the experiential distinction between inner speaking and innerly hearing her own voice. These results highlight the precision and validity of the DES method as a technique of exploring inner experience and the utility of combining such methods with fMRI.

Keywords: inner speech, inner speaking, inner hearing, inner experience, introspection, fMRI, descriptive experience sampling, mind wandering

INTRODUCTION

There is a large literature in both cognitive neuroscience and behavioral psychology that seeks to characterize aspects of inner experience. A growing subset of studies has sought to use fMRI to identify neural correlates of inner experience, for example by studying thinking that is not tightly related to the current task. This has been variously called mind wandering (Kane et al., 2007; Killingsworth and Gilbert, 2010; Smallwood, 2013); inner experience during resting state (Fell, 2013); undirected thought (Klinger and Cox, 1995); stimulus independent thought (Gilbert et al., 2007; Mason et al., 2007); task unrelated thought (Giambra, 1989; Smallwood, 2013); spontaneous thought (Klinger, 2009); daydreaming (Singer, 1966); free thought (Doucet et al., 2012); and so on. Those terms have somewhat different meanings, but all depend centrally on directly apprehended phenomena of inner experience: for example, at time *t* I was thinking about my girlfriend while I was supposed to be doing my homework. Following Christoff (2012), we will refer to all of these as "undirected thoughts."

Capturing the incidence and flow of such inner experience is an intriguing challenge for neuroimaging. Undirected thoughts have been mostly studied either by contrasting passive "baseline" periods with an active, stimulus-driven task (see Shulman et al., 1997) or in terms of connectivity between different brain regions in a resting state (e.g., Greicius et al., 2008). The identification of brain areas and networks associated with rest allows for brain-behavior correlations to be examined: for instance, self-reports of a particular experience, such as anxiety, can be correlated across a group of individuals with responses in particular brain regions (McGuire et al., 1997; Dennis et al., 2011; Doucet et al., 2012). Of particular note, the "default mode network" (DMN) refers to a

set of regions that appear to be consistently anti-correlated with task-positive activity and associated with introspective processes (Raichle and Snyder, 2007; Buckner et al., 2008). There are, however, a range of other networks also thought to be active during resting periods, such as those guiding attentional and executive control (Smallwood, 2013).

Recently a number of novel methods have been developed to induce and examine undirected thoughts, most of which have targeted instances of mind wandering. One approach is to train participants on a particular cognitive task that is conducive to mind wandering, deploy the task in the scanner, ask participants to report instances of mind wandering, and then correlate this with DMN activity (McKiernan et al., 2006; Mason et al., 2007). Another approach has been to ask participants to respond to random probes and report on their subjective state while performing a repetitive task in the scanner (Christoff et al., 2009; Andrews-Hanna et al., 2010; Stawarczyk et al., 2011) or during task-free rest periods (Tusche et al., 2014). For instance, Christoff et al. (2009) probed participants during a go/no-go task and asked them to use Likert scales to indicate (a) whether their attention was focused on the task and (b) whether they were aware of their attentional state at that time ("concurrent awareness"). Self-reported mind wandering states were associated with increased activation of typical areas of the DMN and the central executive network, and this was particularly the case for mind wandering without concurrent awareness (Christoff et al., 2009).

These methods offer novel ways of investigating undirected thoughts. Nevertheless, they all involve a trade-off of some kind. Experiences can be sampled immediately by asking participants to make a forced-choice discrimination while in the scanner. But such techniques almost always preclude detailed descriptions of

experience; experience sampling techniques are typically used simply as classifiers of two or three different "modes" of thinking (such as being "on-task" vs. "off-task"; Christoff et al., 2009). If researchers have wanted to know more about the nature of that state, they have occasionally inquired via questionnaire or interview, but these inquiries have always been retrospective, outside the scanner (Delamillieure et al., 2010; Doucet et al., 2012; Diaz et al., 2013). However, such retrospective reporting undermines the immediacy of the sampling method, because self-reporting on experience after a delay may be subject to a range of reporting biases (Fell, 2013) and is likely to reflect presuppositions about the experience rather than direct apprehensions of experiential phenomena (Hurlburt, 2011).

A technique with the potential to offer an alternative method is descriptive experience sampling (DES; Hurlburt, 1993, 1997, 2011; Hurlburt and Heavey, 2001, 2002, 2006; Hurlburt and Akhter, 2006; Hurlburt and Schwitzgebel, 2007; Heavey and Hurlburt, 2008; Heavey et al., 2012; Hurlburt et al., 2013). DES uses a random beeper to signal participants to attend to the experience that is ongoing at the moment of the beep, and immediately thereafter to jot down notes about that ongoing experience. Within 24 h the DES investigator conducts an "expositional interview" to discover the characteristics of the beeped pristine experience. This process is repeated over a number of days (usually 4–6). DES differs from other experience sampling methods in that it aims to help participants bracket their tendencies to report presuppositions and thereby aims to produce high fidelity descriptions of "pristine" inner experiences—thoughts, feelings, sensations, seeings, hearings, and so on as they naturally occur in a person's everyday environment. It therefore does not specify characteristics that are to be rated in a forced-choice manner, but instead explores characteristics that emerge (Hurlburt, 2009) from the individual person's own experience, untainted (as much as possible) by the investigator's predilections.

There are five experiential phenomena that DES claims occur frequently (Heavey and Hurlburt, 2008); we will call those the 5FP for "5 frequent phenomena": inner speaking (Hurlburt et al., 2013), inner seeing (aka visual imagery), unsymbolized thinking (Hurlburt and Akhter, 2008), sensory awareness (Hurlburt et al., 2009), and feelings (Heavey et al., 2012). However, DES is fundamentally an idiographic procedure, which implies that in any particular individual it is possible that one or more of those characteristics are present, that none of those characteristics are present, or that one or another of those characteristic might be present in an idiosyncratic way. Thus the 5FP are sometimes convenient ways to characterize an individual, sometimes not. That is, DES in general has no expectations about what might emerge as the characteristics of any individual's inner experience. The DES procedure is "open-beginninged" (Hurlburt and Heavey, 2006; Hurlburt and Schwitzgebel, 2007; Hurlburt, 2011) in the sense that the DES interviewer does *not* set out to inquire whether a participant is innerly speaking, or is innerly seeing, or so on. If a participant, describing her experience without benefit of the DES categories, describes (for example) inner speaking (regardless of the words she chooses in so describing), the DES investigator will continue to explore whether inner speaking was ongoing at the moment of the beep.

The DES process can produce surprising results. For example, inner speech is held by some to occur during every waking moment (Archer, 2003; Baars, 2003; Ihde, 2012) and by others to occur about three-quarters of the time (e.g., Klinger and Cox, 1995). DES studies by Hurlburt and colleagues, however, hold that inner speaking is present on average only about one-quarter of the time and in many people never or only very rarely (Heavey and Hurlburt, 2008; Hurlburt et al., 2013). Hurlburt (2011) has argued that discrepancies in reports of many experiential phenomena are common because many if not most people do not know the characteristics of their own inner experience, and instead rely on presuppositions that interfere with their ability to report reliably and accurately. Furthermore, Heavey and Hurlburt (2008) report huge individual differences in the frequency of inner seeing, feelings, and other forms of inner experience, raising further concerns about how assumptions about the nature of inner experience might influence self-report.

In contrast to other methods, then, DES aims to clear out these assumptions and train participants to be more careful in reporting their experiences by avoiding generalizations, focusing on a precise moment (just before the beep), and interviewing participants not just once but on multiple occasions, in an iterative manner (Hurlburt, 2009, 2011). It remains to be seen whether such an idiographic procedure can be of value to neuroscientific investigations of inner experience and undirected thought. The challenge is to convert such a highly detailed method for use with fMRI, and then to test (using evidence from cognitive neuroscience) the claim that DES can offer a reliable reflection of participants' experiences.

Two of us (SK, a neuroscientist, and CF, a psychologist) invited DES-creator Hurlburt (hereafter called RH, whose work we knew but with whom we had no relationship) to put the DES method to the test. We recruited five participants, all unknown to RH and unrelated to us, and invited RH to perform a typical DES investigation with each. Data on all five participants is reported elsewhere as part of a wider study (Hurlburt et al., in preparation). Our aim here is to establish proof-of-principle in a single participant that idiographic investigation of inner experience, such as is provided by DES, can be successfully combined with fMRI. As such we only report here on one of those participants, "Lara," an 18 year-old-woman. As with the other participants, RH trained Lara in the usual DES way (four sampling days in her natural environments); then we delivered random DES beeps to her in nine sessions while she was in an MRI scanner, four random beeps per 25-min session. Immediately after each session, RH conducted a typical DES interview with her about the four beeped experiences. Before the fMRI data were analyzed, RH descriptively characterized those 36 in-the-scanner moments of Lara's experience; these 36 DES characterizations were used to form participant-specific categories on which the fMRI contrasts would be based.

In a preliminary phase at the start of the study, for possible comparison with the DES results, we also asked Lara to perform conventional neuroscience imagination tasks in the scanner: we directed her to generate specific verbal, auditory, visual, emotional, and somatosensory imagery when instructed by prompts such as "to see a pencil" or "to say 'lamp.'" These conventional tasks could then be used to test whether the neuroimaging results derived

from the DES method were plausible, localizable, and comparable to brain activity determined by conventional (non-introspective) means.

Because of the idiographic, open-beginninged nature of DES, at the outset of Lara's DES sampling, we had no expectations about whether one or more of the 5FP might emerge as a salient characteristic of her inner experience. Nonetheless, to facilitate nomothetic comparison it is useful to consider the 5FP if they emerge. It turned out that for Lara, according to DES, sensory awareness was the most frequent of the 5FP (27 occasions or 75% of Lara's 36 in-scanner samples); however, its varying modality (visual, bodily, auditory, etc.) made it an unlikely candidate for neural correlation. Inner seeing occurred in 8 (22%) of Lara's 36 in-scanner samples; however, inner seeing had never occurred in Lara's natural environment DES sampling, so it seemed likely to be an artifact of the scanner situation. Inner speaking occurred in 8 (22%) of Lara's 36 in-scanner samples; it had also occurred in 13% of Lara's natural environment DES sampling, so inner speaking seems a good candidate for further consideration. Of the remaining 5FP phenomena, unsymbolized thinking and feelings were too rare [one occasion (3%) each]. (Percentages do not add to 100% because multiple ratings are possible.) Thus, as a result of Lara's idiographic experiential result, we will focus here on Lara's inner speech-related neural processing.

The neuroimaging literature suggests that language or speech-based samples are typically associated with brain areas such as left inferior frontal gyrus (IFG), superior temporal sulcus (STS), and the superior and middle temporal gyri (Price, 2012; Gernsbacher and Kaschak, 2003), with inner speech processes in particular being linked to activation in left IFG and lateral temporal areas (Shergill et al., 2001, 2002; Jones and Fernyhough, 2007; Marvel and Desmond, 2012; Kühn et al., 2013). Damage to left IFG is also associated with impairment on inner speech tasks (e.g., Geva et al., 2011). If Hurlburt's claims about DES were correct, the fMRI data collected during Lara's eight inner-speaking (according to DES) moments should involve brain activation in some or all of the above-mentioned areas. If this was observed, it would provide support to the principle that DES and fMRI might be profitably combined.

MATERIALS AND METHODS

The study was conducted according to the Declaration of Helsinki, with approval of the German Psychological Society Ethics Committee.

Lara was scheduled for 19 sessions across a 2-week period, which was divided into four phases. She was right-handed and aged 18.

In Phase 1 (*introduction/in-scanner elicitation*), we fully explained the study; administered the initial questionnaires not relevant to the present report; and familiarized Lara with the MRI scanner and procedures. Then Lara entered the scanner, where we conducted a 10 min structural scan and a 5 min resting state scan according to standard fMRI research procedures (keep the eyes closed, stay relaxed and calm). Then we administered the imagination task, derived from a recent fMRI paradigm used by Belardinelli et al. (2009). Participants in the scanner were shown short written prompts to imagine saying (e.g., "to say 'pencil,'" or

"to say 'lamp'"), seeing ("to see a pencil," "to see a lamp"), hearing, feeling, or sensing something. The stimuli were presented in mini-blocks consisting of four prompts of one of the five categories, each shown for 7 s with 1 s inter-stimulus interval (=32 s in total). Thus one mini-block consisted of four seeing prompts; another mini-block consisted of four saying prompts; and so on. Participants were instructed to imagine vividly what was shown on the screen for the duration of the presentation of the prompt. After four prompts a fixation cross was shown for 19 s before the next mini-block of prompts was presented.

In Phase 2 (*natural-environment DES*), we instructed Lara in the use of the DES beeper and the sampling task (Hurlburt and Heavey, 2006; Hurlburt, 2011): she was to wear the beeper in her natural environment for approximately 3 h, during which she would hear (through an earphone) six randomly occurring 700 Hz beeps. Her task was to terminate the beep (a button press) and then immediately to jot down notes about her ongoing inner experience that was "in flight" at the moment the beep sounded. Later that day or the next day she returned for a DES expositional interview about those six beeped experiences; this interview was conducted by RH and at least one and as many as four additional interviewers (the study was part of a training program), usually including some combination of SK, CF, and BAD. The expositional interviews were "iterative" (Hurlburt, 2009, 2011), designed to provide increasing-across-sampling-days skill in apprehending and describing inner experience. Following this interview, Lara returned to her everyday environment, during which (and on the same day) she responded to six more random beeps. The following day she returned for a second expositional interview about the second-sampling-day's six beeped experiences. This sequence was repeated twice more, so that Lara sampled in four natural-environment periods, each followed by an expositional interview.

In Phase 3 (*in-scanner DES*), Lara (having been trained in DES in the natural environment) entered the scanner for a 25-min session with the instruction to keep the eyes open, stay relaxed and calm. That is, we did not set any particular task for her other than to respond to the DES beep (700 Hz, except these beeps were of 1.4 s duration); she had been instructed to note her experience that was ongoing *just prior* to the beep (that is, in the usual DES way). At four quasi-random times, she received a DES beep through a headphone. Immediately after each beep, she jotted a few notes about her experience on a clipboard positioned on her lap (viewable through a mirror); that is, this procedure mirrored as closely as possible the natural-environment DES procedure. Immediately after she exited the scanner she participated in a DES expositional interview (conducted by RH and some combination of SK and BAD) about each of her four randomly beeped experiences, in the order in which they appeared (although doubling back and looking ahead was allowed). This 25-min fMRI scan/four beeps with jotted notes/expositional interview sequence was repeated a total of nine times, resulting in $4 \times 9 = 36$ random samples of experience occurring in $25 \times 9 = 225$ min of fMRI scanning.

In Phase 4 (*post-DES resting state*), Lara entered the scanner for another 10 min structural scan and a 5 min standard resting-state scan. Immediately after exiting the scanner, she completed questionnaires not relevant here. Then she was candidly debriefed.

SCANNING PROCEDURE

Images were collected on a 3T Magnetom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany) using a 32-channel radio frequency head coil. Structural images were obtained using a three-dimensional T1-weighted magnetization-prepared gradient-echo sequence (MPRAGE) based on the ADNI protocol (www.adni-info.org) [repetition time (TR) = 2500 ms; echo time (TE) = 4.77 ms; TI = 1100 ms, acquisition matrix = 256 × 256 × 176, flip angle = 7; 1 mm × 1 mm × 1 mm voxel size]. Functional images were collected using a T2*-weighted echo planar imaging (EPI) sequence sensitive to blood oxygen level dependent (BOLD) contrast (TR = 2000 ms, TE = 30 ms, image matrix = 64 × 64, FOV = 216 mm, flip angle = 80, voxel size 3 mm × 3 mm × 3 mm, 36 axial slices).

fMRI DATA PRE-PROCESSING AND MAIN ANALYSIS

The fMRI data were analyzed using SPM8 software (Wellcome Department of Cognitive Neurology, London, UK). The first four volumes of all EPI series were excluded from the analysis to allow the magnetization to approach a dynamic equilibrium. Data processing started with slice time correction and realignment of the EPI datasets. A mean image for all EPI volumes was created, to which individual volumes were spatially realigned by means of rigid body transformations. The structural image was co-registered with the mean image of the EPI series. Then the structural image was normalized to the Montreal Neurological Institute (MNI) template, and the normalization parameters were applied to the EPI images to ensure an anatomically informed normalization. A commonly applied filter of 8 mm full-width at half maximum (FWHM) was used. Low-frequency drifts in the time domain were removed by modeling the time series for each voxel by a set of discrete cosine functions to which a cut-off of 128 s was applied. The statistical analyses were performed using the general linear model (GLM).

The imagination task was modeled as blocks with a duration of 32 s. The beeps of the DES procedure were modeled as events on the onset of the beep with a duration of 0. These vectors were convolved with a canonical hemodynamic response function (HRF) and its temporal derivatives.

For the DES procedure, we asked RH, on the basis of the DES expositional interviews he (and others) had conducted, to classify each of Lara's 36 in-the-scanner experiences into four modalities: verbal, visual, bodily, and auditory (categories could overlap). We also asked RH to classify each of Lara's experiences according to which (if any) of the 5FP (inner speaking, inner seeing, unsymbolized thinking, feeling, and sensory awareness) were present. RH's classifications were checked by at least one other person who had been present at the relevant interview; disagreements were resolved by consensus. Then, regressors were built coding the categories that RH had assigned to the 36 events. For display purposes the resulting SPMs were thresholded at $p < 0.001$ and a significant effect was reported when the volume of the cluster was greater than the Monte-Carlo-simulation-determined minimum cluster size above which the probability of type I error was < 0.05 (AlphaSim; Ward, 2000). The resulting maps were overlaid onto a normalized T1 weighted MNI template (colin27)

and the coordinates reported correspond to the MNI coordinate system.

RESULTS

In Phase 1 (*introduction/in-scanner elicitation*) we investigated whether Lara would produce results that aligned with what conventional neuroscience would predict for inner speaking. **Figure 1A** shows a comparison of mini-blocks in which Lara had been instructed to imagine herself speaking (but not actually speaking aloud) compared against mini-blocks in which a fixation cross was shown. This figure shows that Lara's brain produced the predicted activation of the inner speech network: left IFG and STS as well as superior and middle temporal gyrus (see also **Table 1A**).

FIGURE 1 | (A) Contrast of inner saying > fixation in a conventional neuroscientific mini-block design; **(B)** Contrast of verbal modality > all other modalities (visual, bodily, and auditory) as defined by DES; **(C)** Contrast of inner speaking > all other categories as defined by descriptive experience sampling (DES) 5 frequent phenomena (5FP); **(D)** Contrast of inner speaking > inner hearing of idiographic DES categories.

There was no fMRI data collected during Phase 2. In Phase 3 we ask first whether the DES interviews conducted by RH are capable of classifying Lara's verbal experience in ways that correspond to her neurophysiological activation. **Figure 1B** shows that when we compare Lara's brain activity in those moments that RH classified as verbal to those classified as nonverbal (visual, bodily, or auditory), Lara showed the predicted activation of left IFG (see also **Table 1B**).

Next in Phase 3, we examine the 5FP category of inner speaking. Samples that included inner speaking were spread throughout the scanning sessions: three samples occurred during the third scanning session, one each during the fourth, seventh, and eighth scans, and two during the ninth scan. We modeled brain activity across all 36 samples as a function of whether RH had said inner speaking was or was not present. We did the same kind of univariate modeling across all 36 samples for each of the remaining four 5FP characteristics (that is, for inner seeing, for unsymbolized thinking, for feelings, and for sensory awareness). Then across the eight samples in which RH had said inner speaking was present, we compared the average of results of the inner speaking model to the average of the results of the four remaining models; this

Table 1A | Conventional imagination task: inner saying > fixation (FWE p < 0.05).

Area	BA	Peak coordinates (MNI)	t-score	Extent	p
Left rolandic operculum	6	−48, 2, 46	11.74	430	0.000
Left dorsolateral prefrontal cortex (DLPFC), inferior frontal gyrus (IFG)	46, 45	−39, 35, 34	9.14	337	0.000
Left middle temporal gyrus	21	−51, −49, 10	8.92	80	0.000
Left middle temporal gyrus	21	−66, −46, 4	8.06	77	0.000
Right visual cortex	18	30, −97, 4	8.00	78	0.000
Right cerebellum (Crus 2)		18, −82, −32	7.97	272	0.000
Left visual cortex	18	−33, −97, −5	7.62	23	0.000
Left angular gyrus	39	−45, −61, 49	7.02	107	0.000
Right temporal pole	38	51, 20, −29	6.60	103	0.000
Left temporal pole	38	−33, 20, −26	6.60	106	0.000
Ventromedial prefrontal cortex	11	−3, 53, −17	6.04	26	0.000
Left IFG	45	−51, 20, 1	5.98	25	0.000

Table 1B | Descriptive experience sampling modalities: verbal modality > all other categories (p < 0.001, k > 22).

Area	BA	Peak coordinates (MNI)	t-score	Extent	p
Right postcentral gyrus	3	48, −22, 43	5.05	102	0.000
Left IFG	45	−51, 29, 10	4.24	31	0.000
Anterior cingulate gyrus	32	0, 20, 43	4.02	34	0.000

analysis indicated the presence of activity in left IFG, the core of the inner speech network (**Figure 1C**, **Table 1C**). We compared inner speaking to the non-inner-speaking samples instead of comparing against baseline because baseline includes times during which Lara was responding to (jotting down notes about) samples.

Because DES is primarily an idiographic technique, it is held to be capable of describing characteristics that apply to one individual, regardless of whether those characteristics are important for many or any other individuals. Therefore we asked RH to identify characteristics that might emerge from Lara's participation in the DES procedure (during either or both the natural environment and the in-the-scanner phases) that are not standard modality features (verbal, visual, bodily, and auditory) and that are not identified as 5FP, regardless of whether the feature was a characteristic of any other DES participant. One such feature that RH noted was that when Lara experienced inner words, they ranged on a continuum from innerly spoken to innerly heard (Hurlburt et al., 2013). The distinction between speaking and hearing may be illustrated by the metaphor of speaking into a tape recorder (production) and hearing your voice being played back (reception). Contrasting Lara's brain activity during moments of inner speaking vs. moments of inner hearing of her own voice resulted in increased activity in left IFG (**Figure 1D**, **Table 1D**).

DISCUSSION

To summarize, whether prompted using a conventional imagery-based fMRI paradigm or classified via use of the DES, Lara's

Table 1C | Descriptive experience sampling 5FP: inner speaking > all other categories (p < 0.001, k > 22).

Area	BA	Peak coordinates (MNI)	t-score	Extent	p
Left IFG	45	−36, 32, 19	4.96	88	0.000
Left DLPFC	9	−12, 41, 43	4.26	58	0.000

Table 1D | Descriptive experience sampling idiographic: inner speaking > Inner hearing (p < 0.001, k > 22).

Area	BA	Peak coordinates (MNI)	t-score	Extent	p
Left IFG	45	−51, 41, −5	3.87	23	0.000

experiences of inner speaking were, as expected, reliably associated with activation in left IFG. This validates the suggestion that it is indeed possible for the DES procedure to apprehend features of inner experience and to do so as they naturally occur—the present study, for example, did not set out to investigate inner speaking; it set out to investigate naturally occurring characteristics of Lara's inner experience *whatever those characteristics might be*. DES identified inner speaking as an important feature of Lara's experience, and our study design allowed us then to demonstrate predicted fMRI activations during the occurrence of that feature.

The validation of high fidelity apprehensions of inner experience demonstrated in the present study should not be taken as a validation of all introspective or subjective reports—DES is, by Nisbett and Wilson's (1977; cf. Hurlburt, 2011, p. 195) analysis, an exceptional method. Furthermore, the present study should not be taken as a validation of all DES-type reports—the demonstration here was by just one DES investigator (and his colleagues) with one participant. However, Hurlburt and Heavey (2002) showed that inter-rater reliability could be high between DES practitioners. A related question is the one raised previously, namely why we presented the case of Lara and not some or all of the other participants. First, it is not possible to combine disparate idiographic cases in a single journal article. Second, even if the case of Lara were the only interesting case of the five, it is enough to establish an important principle: fMRI data and particular-moment-experiential data can be profitably combined at least for some participants in some situations.

The combination of phenomenology and neurophysiology might be understood as a validation of DES: claims about private experience are always questionable, and the fact that the DES claims correlate with known neurophysiological results lends nontrivial support to the adequacy of the DES claims. However, the validation could be understood to operate the other way around: that the DES descriptions lend support to fMRI techniques as a way of investigating short-duration phenomenological characteristics. Either way, such a merger might answer important questions that are impossible even to pose without experiential data aimed at particular moments of consciousness. Here are two examples. First, Lara's inner speaking results show that (at least for Lara) there is a phenomenological distinction between innerly speaking one's voice and innerly hearing one's voice being spoken. To our knowledge, such a distinction is typically not attended to in contemporary models of inner speech (Fernyhough, 2004; Scott, 2013), but it may be a crucial one for future studies. For example, theories of auditory verbal hallucinations (AVHs) that emphasize monitoring of inner speech (e.g., Frith, 1995) may benefit from investigation of localizable differences between inner speaking and inner hearing.

Also of interest was the relatively wider spread of activation associated with imagined inner speech (**Figure 1A**) as compared to unprompted moments of inner speaking classified by the DES (**Figures 1B,D**). It must be borne in mind that these activation maps have differing levels of temporal precision in that the imagery data come from a block design (continually producing bits of inner speech over an extended period of time), whereas the DES samples, by their nature, are targeted at very specific moments.

Nevertheless, one could speculate that these results reflect genuine differences—that is, when Lara is prompted to imagine speaking, her *actual* inner speaking is somewhat different (possibly dramatically different) from her naturally occurring inner speaking both phenomenologically and neurologically (see Hurlburt et al., in preparation). One interpretation is that Lara's actual experience following the inner speech prompt also included some processing of the prompt itself and some monitoring of inner speech production. It should be noted that prompted inner speech is an unnatural phenomenon (Jones and Fernyhough, 2007), occurring very rarely outside of psychological research laboratories. However, nearly all psychological studies of inner speech are either retrospective or of the prompted variety.

Given that this is a single case, we do not know whether distinctions such as between inner speaking and inner hearing and between prompted and unprompted inner speaking reflect idiosyncratic characteristics of Lara (and/or of RH) or are characteristic of other individuals who would regularly report inner speaking as part of their everyday experience. The present study cannot answer such questions, but it does provide a method whereby such questions might be answered.

Furthermore, this study does not by itself establish a principle about introspections in general, because it investigated only one method (DES) and one investigator (RH). For example, this study should *not* be understood as saying that we should simply believe people when they tell us they are talking to themselves; Hurlburt (2011) holds that people are often substantially mistaken about such reports unless an adequate method is used. This study should *not* be understood as saying that questionnaires about experience or non-DES experience sampling are valid descriptors of experience (Hurlburt and Heavey, 2014; Alderson-Day et al., in preparation; Hurlburt et al., in preparation). This study does not explore the boundaries or parameters of confidence in DES (or in RH)—that is, it does not characterize the situations where we can be more (or less) confident about the correspondence between self-reports and associated brain activity.

However, this study does suggest a new set of opportunities for cognitive neuroscience investigations. Most fMRI studies ask the participant in the scanner to perform a task that indirectly invokes a particular set of brain functions in the scanner. Whether receptive (e.g., merely viewing a flashing display) or active (e.g., memorizing syllables), the aim of those tasks is indirect in the sense that the task and stimuli are presumed to elicit the desired brain functions. That is, participants do not observe or report any aspect of their brain or mental function; they simply engage in the task that presumably indirectly evokes the brain function.

As noted in the Introduction, some fMRI studies ask participants in the scanner directly to rate their cognition or mental activity on some predefined measure (e.g., Christoff et al., 2009). But until now, no study has tried to link fMRI measurements to naturally occurring rather than task-elicited features of a participant's experiential phenomena in the scanner (Hurlburt et al., in preparation). Now that we have established that such studies are possible, future investigations can explore the utility, limitations, and boundaries of such studies, for example comparing DES with other introspection methods and their correlation with brain activity.

REFERENCES

Andrews-Hanna, J., Reidler, J. S., Huang, C., and Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.* 104, 322–335. doi: 10.1152/jn.00830.2009

Archer, M. S. (2003). *Structure, Agency and the Internal Conversation.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139087315

Baars, B. (2003). How brain reveals mind neural studies support the fundamental role of conscious experience. *J. Conscious. Stud.* 10, 100–114.

Belardinelli, O. M., Palmiero, M., Sestieri, C., Nardo, D., Di Matteo, R., Londei, A., et al. (2009). An fMRI investigation on image generation in different sensory modalities: the influence of vividness. *Acta Psychol. (Amst.)* 132, 190–200. doi: 10.1016/j.actpsy.2009.06.009

Buckner, R. L., Andrews-Hanna, J., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Christoff, K. (2012). Undirected thought: neural determinants and correlates. *Brain Res.* 1428, 51–59. doi: 10.1016/j.brainres.2011.09.060

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., and Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8719–8724. doi: 10.1073/pnas.0900234106

Delamillieure, P., Doucet, G., Mazoyer, B., Turbelina, M., Delcroix, N., Mellet, E., et al. (2010). The resting state questionnaire: an introspective questionnaire for evaluation of inner experience during the conscious resting state. *Brain Res. Bull.* 81, 565–573. doi: 10.1016/j.brainresbull.2009.11.014

Dennis, E. L., Gotlib, I. H., Thompson, P. M., and Thomason, M. E. (2011). Anxiety modulates insula recruitment in resting-state functional magnetic resonance imaging in youth and adults. *Brain Connect.* 1, 245–254. doi: 10.1089/brain.2011.0030

Diaz, B. A., VanDerSluis, S., Moens, S., Benjamins, J. S., Migliorati, F., Stoffers, D., et al. (2013). The Amsterdam resting state questionnaire reveals multiple phenotypes of resting-state cognition. *Front. Hum. Neurosci.* 7:446. doi: 10.3389/fnhum.2013.00446

Doucet, G., Naveau, M., Petit, L., Zago, L., Crivello, F., Jobard, G., et al. (2012). Patterns of hemodynamic low-frequency oscillations in the brain are modulated by the nature of free thought during rest. *Neuroimage* 59, 3194–3200. doi: 10.1016/j.neuroimage.2011.11.059

Fell, J. (2013). Unraveling inner experiences during resting state. *Front. Hum. Neurosci.* 7:409. doi: 10.3389/fnhum.2013.00409

Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas Psychol.* 22, 49–68. doi: 10.1016/j.newideapsych.2004.09.001

Frith, C. D. (1995). *The Cognitive Neuropsychology of Schizophrenia.* New York: Psychology Press.

Gernsbacher, M. A., and Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annu. Rev. Psychol.* 54, 91–114. doi: 10.1146/annurev.psych.54.101601.145128

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., and Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain* 134, 3071–3082. doi: 10.1093/brain/awr232

Giambra, L. M. (1989). Task-unrelated thought frequency as a function of age: a laboratory study. *Psychol. Aging* 4, 136–143. doi: 10.1037/0882-7974.4.2.136

Gilbert, S. J., Dumontheil, I., Simons, J. S., Frith, C. D., and Burgess, P. W. (2007). Comment on "wandering minds: the default network and stimulus-independent thought." *Science* 317:43. doi: 10.1126/science.1140801

Greicius, M. D., Supekar, K., Menon, V., and Dougherty, R. F. (2008). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* 19, 72–78. doi: 10.1093/cercor/bhn059

Heavey, C. L., and Hurlburt, R. T. (2008). The phenomena of inner experience. *Conscious. Cogn.* 17, 798–810. doi: 10.1016/j.concog.2007.12.006

Heavey, C. L., Hurlburt, R. T., and Lefforge, N. L. (2012). Toward a phenomenology of feelings. *Emotion* 12, 763–777. doi: 10.1037/a0026905

Hurlburt, R. T. (1993). *Sampling Inner Experience in Disturbed Affect.* New York: Plenum Press. doi: 10.1007/978-1-4899-1222-0

Hurlburt, R. T. (1997). Randomly sampling thinking in the natural environment. *J. Consult. Clin. Psychol.* 65, 941–949. doi: 10.1037/0022-006X.65.6.941

Hurlburt, R. T. (2009). Iteratively apprehending pristine experience. *J. Conscious. Stud.* 16, 156–188.

Hurlburt, R. T. (2011). *Investigating Pristine Inner Experience.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511842627

Hurlburt, R. T., and Akhter, S. A. (2006). The descriptive experience sampling method. *Phenom. Cogn. Sci.* 5, 271–301. doi: 10.1007/s11097-006-9024-0

Hurlburt, R. T., and Akhter, S. A. (2008). Unsymbolized thinking. *Conscious. Cogn.* 17, 1364–1374. doi: 10.1016/j.concog.2008.03.021

Hurlburt, R. T., and Heavey, C. L. (2001). Telling what we know: describing inner experience. *Trends Cogn. Sci.* 5, 400–403. doi: 10.1016/S1364-6613(00)01724-1

Hurlburt, R. T., and Heavey, C. L. (2002). Interobserver reliability of descriptive experience sampling. *Cogn. Ther. Res.* 26, 135–142. doi: 10.1023/A:1013802006827

Hurlburt, R. T., and Heavey, C. L. (2006). *Exploring Inner Experience: The Descriptive Experience Sampling Method.* Amsterdam/Philadelphia: John Benjamins Publishing Company. doi: 10.1075/aicr.64

Hurlburt, R. T., and Heavey, C. L. (2014). Investigating pristine inner experience: implications for experience sampling and questionnaires. *Conscious. Cogn.* (in press).

Hurlburt, R. T., Heavey, C. L., and Bensaheb, A. (2009). Sensory awareness. *J. Conscious. Stud.* 16, 231–251.

Hurlburt, R. T., Heavey, C. L., and Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Conscious. Cogn.* 22, 1477–1494. doi: 10.1016/j.concog.2013.10.003

Hurlburt, R. T., and Schwitzgebel, E. (2007). *Describing Inner Experience?* Cambridge, MA: MIT Press.

Ihde, D. (2012). *Listening and Voice.* New York: SUNY Press.

Jones, S. R., and Fernyhough, C. (2007). Neural correlates of inner speech and auditory verbal hallucinations: a critical review and theoretical integration. *Clin. Psychol. Rev.* 27, 140–154. doi: 10.1016/j.cpr.2006.10.001

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., and Kwapil, T. R. (2007). For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. *Psychol. Sci.* 18, 614–621. doi: 10.1111/j.1467-9280.2007.01948.x

Killingsworth, M. A., and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science* 330, 932–932. doi: 10.1126/science.1192439

Klinger, E. (2009). *Daydreaming and Fantasizing: Thought Flow and Motivation.* New York: Psychology Press.

Klinger, E., and Cox, W. M. (1995). Dimensions of thought flow in everyday life. *Imagin. Cogn. Pers.* 7, 105–128. doi: 10.2190/7K24-G343-MTQW-115V

Kühn, S., Schmiedek, F., Brose, A., Schott, B. H., Lindenberger, U., and Lövden, M. (2013). The neural representation of intrusive thoughts. *Soc. Cogn. Affect. Neurosci.* 8, 688–693. doi: 10.1093/scan/nss047

Marvel, C. L., and Desmond, J. E. (2012). From storage to manipulation: how the neural correlates of verbal working memory reflect varying demands on inner speech. *Brain Lang.* 120, 42–51. doi: 10.1016/j.bandl.2011.08.005

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., and Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science* 315, 393–395. doi: 10.1126/science.1131295

McGuire, P. K., Robertson, D., Thacker, A., David, A. S., Kitson, N., Frackowiak, R. S. J., et al. (1997). Neural correlates of thinking in sign language. *Neuroreport* 8, 695–698. doi: 10.1097/00001756-199702100-00023

McKiernan, K. A., D'Angelo, B. R., Kaufman, J. N., and Binder, J. R. (2006). Interrupting the "stream of consciousness": an fMRI investigation. *Neuroimage* 29, 1185–1191. doi: 10.1016/j.neuroimage.2005.09.030

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84, 231–259. doi: 10.1037/0033-295X.84.3.231

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847. doi: 10.1016/j.neuroimage.2012.04.062

Raichle, M. E., and Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage* 37, 1083–1090. doi: 10.1016/j.neuroimage.2007.02.041

Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychol. Sci.* 24, 1824–1830. doi: 10.1177/0956797613478614

Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R. M., et al. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Hum. Brain Mapp.* 16, 219–227. doi: 10.1002/hbm. 10046

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C., Murray, R. M., and McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychol. Med.* 31, 241–253. doi: 10.1017/S003329170100335X

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *J. Cogn. Neurosci.* 9, 648–663. doi: 10.1093/cercor/7.3.193

Singer, J. L. (1966). *Daydreaming.* New York: Crown.

Smallwood, J. (2013). Not all minds that wander are lost: the importance of a balanced perspective on the mind wandering state. *Front. Psychol.* 4:441. doi: 10.3389/fpsyg.2013.00441

Stawarczyk, D., Majerus, S., Maquet, P., and D'Argembeau, A. (2011). Neural correlates of ongoing conscious experience: both task-unrelatedness and stimulus-independence are related to default network activity. *PLoS ONE* 6:e16997. doi: 10.1371/journal.pone.0016997

Tusche, A., Smallwood, J., Bernhardt, B. C., and Singer, T. (2014). Classifying the wandering mind: revealing the affective content of thoughts during task-free rest periods. *Neuroimage* 97, 107–116. doi: 10.1016/j.neuroimage.2014.03.076

Ward, B. D. (2000). *Simultaneous Inference for FMRI Data. Afni Alphasim Documentation.* Milwaukee, WI: Medical College of Wisconsin.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Relations among questionnaire and experience sampling measures of inner speech: a smartphone app study

Ben Alderson-Day* and Charles Fernyhough

Department of Psychology, Durham University, Durham, UK

Inner speech is often reported to be a common and central part of inner experience, but its true prevalence is unclear. Many questionnaire-based measures appear to lack convergent validity and it has been claimed that they overestimate inner speech in comparison to experience sampling methods (which involve collecting data at random timepoints). The present study compared self-reporting of inner speech collected via a general questionnaire and experience sampling, using data from a custom-made smartphone app (*Inner Life*). Fifty-one university students completed a generalized self-report measure of inner speech (the Varieties of Inner Speech Questionnaire, VISQ) and responded to at least seven random alerts to report on incidences of inner speech over a 2-week period. Correlations and pairwise comparisons were used to compare generalized endorsements and randomly sampled scores for each VISQ subscale. Significant correlations were observed between general and randomly sampled measures for only two of the four VISQ subscales, and endorsements of inner speech with evaluative or motivational characteristics did not correlate at all across different measures. Endorsement of inner speech items was significantly lower for random sampling compared to generalized self-report, for all VISQ subscales. Exploratory analysis indicated that specific inner speech characteristics were also related to anxiety and future-oriented thinking.

Keywords: covert speech, dialog, introspection, verbal thinking, self-talk

Introduction

"*Human beings talk to themselves every moment of the waking day*" (Baars, 2003, p. 106)

Inner speech—talking to oneself silently and internally—seems to be a central part of conscious experience. Cognitive and developmental research on inner speech has led to it being associated with a variety of activities and skills, including problem-solving, memory, and self-reflection (Sokolov, 1975; Morin, 2005; Perrone-Bertolotti et al., 2014). Less information exists, however, on the extent and nature of everyday inner speech use, in part because of methodological difficulties in measuring the phenomenon reliably.

Studies seeking to empirically investigate inner speech have tended to use generalized self-report methods, such as "thought-listing," diaries, or questionnaires. For example, Morin et al. (2011) asked a sample of 380 university students to list, in an open format, "as many verbalizations as

they typically address to themselves." They then coded responses for their content (such as positive or negative statements, or comments about other people) and their apparent function (such as self-regulation). By far the most common kind of content reported was self-talk about oneself, including statements relating to self-evaluation and utterances concerning emotions, personal relationships, and physical appearance. In terms of functions, inner speech was most commonly used for planning ahead, remembering previous events, and motivating behavior. These findings were consistent with a diary-based study of future-thinking by D'Argembeau et al. (2011): participants were asked to list a selection of thoughts each day, and rate their content and phenomenological qualities, including whether they were in inner speech or in other modalities (such as visual imagery). Compared to other modalities, inner speech was particularly associated with planning and decision-making. In addition, inner speech was more likely to occur for negative or neutral thoughts than positive thoughts.

Broadly similar results have been provided by questionnaire-based studies of inner speech and private speech (i.e., speech which is external but self-directed). Duncan and Cheyne (1999) developed a questionnaire to study self-verbalizations (a mixture of private and inner speech) and examined the most common factors that arose in students. Overall, verbalizations were most often used for "cognitive-attentional" activities (i.e., trying to remember something, or avoid distractions) and organizing behavior, such as planning out a series of actions. Brinthaupt et al.'s (2009) Self-Talk Scale (STS) is a measure that emphasizes inner speech more specifically, and includes four main factors: social assessment, self-reinforcement, self-criticism, and self-management. Across a series of experiments, they showed that individual differences in self-talk frequency related to various cognitive, behavioral, and mood factors. For instance, those who reported more frequent instances of critical or evaluative self-talk also reported lower self-esteem and a greater number of automatic negative thoughts.

Most recently, McCarthy-Jones and Fernyhough (2011) devised the Varieties of Inner Speech Questionnaire (VISQ). The VISQ asks about variations in inner speech that reflect its putative origin in external communication (Vygotsky, 1987), whereby silent or covert self-talk reflects an internalized and transformed version of outer dialog and social interaction. Thus, it asks about inner speech that varies in structure (such as being in full sentences, or single words, or having a turn-taking quality), identity (such as involving interaction with other people), and, consistent with other scales, self-regulatory behaviors (such as encouraging or criticizing oneself). McCarthy-Jones and Fernyhough's (2011) initial use of the scale displayed a four-factor structure, encompassing *dialogic* inner speech (inner speech with a conversational quality), *evaluative/motivational* inner speech (i.e., self-regulatory inner speech), *other people* in inner speech (e.g., comments from other agents, such as relatives), and *condensed* inner speech (the extent to which inner speech was abbreviated in some way). They found that dialog-like and evaluative characteristics of inner speech were very common (endorsed by 75–80% of participants), while the presence of other people in

inner speech and condensed inner speech also appeared in a substantial minority of individuals. Moreover, the different characteristics of inner speech were specifically related to differing aspects of psychopathology: for instance, McCarthy-Jones and Fernyhough (2011) observed that inner speech containing other people or evaluative characteristics was related to self-reported scores for anxiety, while a separate study by Alderson-Day et al. (2014) found that evaluative inner speech was also associated with lower self-esteem. Taken together, the above studies point to an experience of everyday inner speech that often involves self-evaluation and thinking about the future, but can sometimes be associated with low mood and feelings of anxiety.

However, the validity of investigating inner speech in this way—that is, relying on participants' generalized self-reports—has been questioned. Using the same sample as Morin et al. (2011), Uttl et al. (2011) assessed the reliability and validity statistics of their own self-report method, along with Duncan and Cheyne's (1999) Self-Verbalization Questionnaire, Brinthaupt et al.'s (2009) STS, and two other self-talk measures: the Inner Speech Scale (ISS; Siegrist, 1995) and the Self-Talk Inventory (Calvete et al., 2005). While all of the scales showed good internal reliability, they showed very little convergent validity (with only the STS and ISS being more than weakly correlated) and tended not to correlate at all with the inner speech reports collected by Morin et al. (2011). That is, what people freely list as generally occurring in their inner speech, and how they respond to various general questionnaires about the same topic, did not seem to closely correspond.

A second, related cause for concern is the possibility that individuals may over-estimate their own inner speech when they are asked about it in generalized terms. While some older studies have claimed very high frequency rates for inner speech, i.e., >50% of daily samples (Klinger and Cox, 1987; Goldstein and Kenen, 1988), it has recently been suggested that inner speech is far from ubiquitous and universal. Based on their studies with Descriptive Experience Sampling—a method where participants are prompted at random by a beeper to record their everyday inner experience, and are then interviewed about that experience in depth—Heavey and Hurlburt (2008) argued that inner speech only occurs in 26% of random samples. Hurlburt et al. (2013) suggest that inner speech questionnaires over-estimate the occurrence of inner speech because they ask participants to provide a general estimate of its occurrence, which is more likely to reflect participants' preconceptions about their inner speech rather than its actual presence on a moment-by-moment basis. Hurlburt et al.'s (2013) view is supported by evidence from ecological momentary assessment studies, which have often reported over-estimation of traits and behaviors when they are gathered via generalized self-report compared to random or momentary sampling (see Shiffman et al., 2008, for a review).

As has been noted elsewhere (McCarthy-Jones and Fernyhough, 2011; Morin et al., 2011), new and alternative methods are needed to fully assess whether generalized self-report measures of inner speech can be reliable and valid indicators of everyday inner speech (Alderson-Day and Fernyhough, 2014). One way to do that is to combine questionnaire methods with

experience sampling (Csikszentmihalyi and Larson, 1987), in which participants are prompted at random intervals to provide data. DES as used by Hurlburt et al. (2013) is one kind of in-depth experience sampling, but it is very resource-intensive and typically only used in small groups of participants. An alternative option is to use smartphone-based assessments to gather large amounts of randomly sampled data. Experience sampling via iPhones and other devices has previously been used successfully to examine the relations between mood and mind-wandering (Killingsworth and Gilbert, 2010). In some cases, such methods have also been reported to change participants' self-reporting skills (a phenomenon known as 'reactivity'): for instance, Runyan et al. (2013) observed increased levels of self-awareness and better time management in a group of university students who were assigned to use an experience sampling app (iHabit) compared to controls who simply completed general questionnaires.

We applied the smartphone app methodology to the study of inner speech using a custom-made app, *Inner Life*. Inner Life works by prompting participants at random, twice a day, to answer a short series of questions about their ongoing thoughts, feelings, and behavior. Specifically, volunteers are asked to indicate what they were doing immediately prior to noticing the alert from the app. In the present study they did this for 2 weeks, building up a maximum of 28 data points for various aspects of inner life, including inner speech, activity, mood, and mind-wandering.

In addition, we asked participants to complete a generalized measure of inner speech—the VISQ (McCarthy-Jones and Fernyhough, 2011)—at the start and end of the two-week period. The VISQ was chosen for the following reasons: (1) it is unique among questionnaire measures in asking specifically about inner speech, rather than more general "self-talk" (which could include overt and covert verbalizations); (2) it covers a broad range of phenomenological features of inner speech, with relevance to form, function, and identity; and (3) its subscales show good internal reliability and reasonable validity, in terms of correlations with other self-report traits (such as proneness to anxiety).

We hypothesized that (1) generalized endorsements of inner speech characteristics would be reliable indicators of inner speech incidence recorded via random sampling, but that (2) moment-by-moment endorsements of inner speech would, on average, be lower than generalized endorsements of inner speech, based on the claims of Hurlburt et al. (2013).

We also set out to explore how inner speech related to other measures of mood and thinking identified in previous studies on the topic. First, based on prior links between inner speech, self-reflection, anxiety, and positive and negative thinking (D'Argembeau et al., 2011; McCarthy-Jones and Fernyhough, 2011; Morin et al., 2011; Alderson-Day et al., 2014), we included mean scores for happiness and anxiety in this analysis. Second, we assessed relations between inner speech and temporal thinking (i.e., whether or not participants were generally thinking about the past, present, or future) based on the apparent involvement of inner speech in remembering the past and planning/future thinking (D'Argembeau et al., 2011). If inner speech served the apparent temporality-related functions reported by D'Argembeau et al. (2011) and Morin et al. (2011), then different characteristics of the

VISQ could be expected to be associated with temporal thinking about the past and future.

Materials and Methods

Participants
Fifty-one university students (38 female; age M = 19.88, SD = 2.96) were recruited via a participant-pool advertisement. All participants had English as their first language. Participation was rewarded with course credit. All procedures were approved by a local university ethics committee.

Procedure
Participants were provided with a link to download Inner Life via an online information and consent page. Inner Life was made available to participants for Android phones (via a private link) or iOS (via the *Testflight* app-testing service). The app download page included instructions on how to use the app and respond to each alert. Participants were encouraged to respond to the app as soon as it was safe to do so, and to answer based on what was happening immediately prior to the moment they noticed the alert.

When the app was first opened, participants were prompted to complete a battery of generalized questionnaire measures, taking roughly 5–10 min to complete (including the full VISQ). Following this, Inner Life was configured to deliver two alerts a day for 14 days (see **Figure 1**). Each alert contained 12–18 questions about ongoing mental phenomena, taking less than 2 min to complete. The alerts occurred at random intervals within two 3-h windows each day, one early and one late (by default, this was set to 9 am–12 pm and 2 pm–5 pm). Participants could choose when their windows occurred to avoid intrusion. The only limit on window selection was that one had to be before 2 pm, and one after (in order to ensure a spread of responses between morning and afternoon). On the 14th day of testing, the final alert contained a second general self-report battery for assessment of test–retest reliability.

Measures
Inner Life collected a large battery of mood and psychopathology data that are being analyzed as part of a larger ongoing study. Here, we report the main variables associated with inner speech.

Inner Speech via General Questionnaire
The VISQ (McCarthy-Jones and Fernyhough, 2011) was used to assess phenomenological characteristics of inner speech at the start and end of the two-week period. The VISQ is an 18-item scale containing four subscales: *Dialogic Inner Speech* (Dialogic henceforth), *Evaluative/Motivational Inner Speech* (Evaluative), *Other People in Inner Speech* (Other People), and *Condensed Inner Speech* (Condensed). Each item is answered on a 6-point Likert scale ranging from "Certainly does not apply to me" (1) to "Certainly applies to me" (6). Each of the subscales has very good internal reliability (Cronbach's alpha > 0.80) and moderate/good test–retest reliability (> 0.6). Participants completed the full VISQ on entry to the study (T1) and at the end of the study (T2).

FIGURE 1 | Inner Life. Participants received two alerts a day **(A)** to answer a short battery of questions on inner speech and related topics **(B)**. Alerts would occur at random within two 3-hour windows, selected by the participant **(C)**.
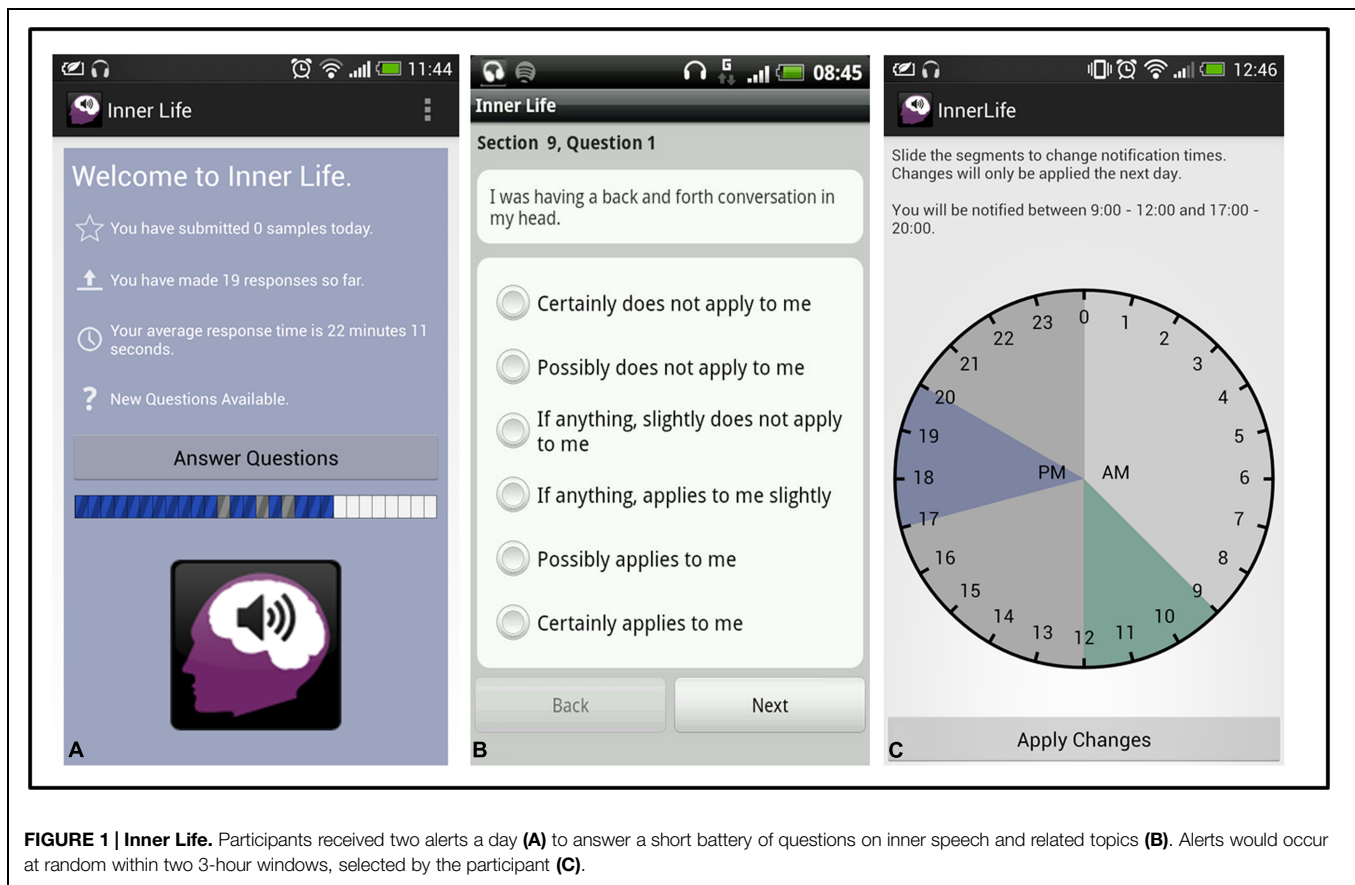
Inner Speech via Random Sampling

Inner speech collected via the random alerts was assessed using four adapted items from the full VISQ.

For each subscale the highest loading item from the original VISQ factor analysis by McCarthy-Jones and Fernyhough (2011) was selected and then reworded to refer to the current moment:

At the time of the alert:
(1) *I was having a back and forth conversation in my head* (Dialogic)
(2) *My thinking was shortened compared to my normal, out-loud speech* (Condensed)
(3) *I was having the experience of other people's voices* (Other People)
(4) *I was evaluating my behavior using my inner speech* (Evaluative)

Each item was presented sequentially, with participants answering using the same 6-point Likert scale as the full VISQ.

Other Measures

A selection of alerts also included ratings for three other relevant variables: happiness, anxiety, and temporal thinking.

- For **happiness** and **anxiety** participants indicated their current mood level on a visual analog scale, from 0–10. 50% of alerts contained happiness questions, while 75% contained anxiety questions. Questions about each were evenly spaced through the 14-day sampling period.
- For **temporal thinking**, participants were asked to indicate whether they were thinking about the past, present, or future at the moment of the alert. Fifty percentage of samples contained a question about temporal thinking. Samples with a temporal thinking question alternated each day (i.e., Day 1 contained a question in the AM window, Day 2 in the PM window).

Analysis

As the large majority of outcome variables were non-normally distributed, non-parametric tests were used (Spearman's Rho and Wilcoxon Signed Ranks Test). To compare generalized endorsements and randomly sampled incidences of inner speech, mean scores were calculated from the random alerts for each VISQ factor and then scaled up to provide "total" scores. For example, a mean score of 4 on the Dialogic item would receive a total score of 16, based on the fact that the general VISQ contains four items in the Dialogic subscale. Bivariate correlations were used to examine reliability of T1 scores. Wilcoxon tests were used to compare overall levels of endorsement for inner speech characteristics. T1 and T2 general VISQ scores were then assessed for test–retest reliability and compared for overall score, to assess changes in reporting following use of the app. To control for multiple comparisons for the four

VISQ subscales, alpha was adjusted to $p < 0.0125$ (i.e., 0.05/4), while p-values between 0.0125 and 0.05 were treated as trends (see "Overall Characteristics of Inner Speech," "Similarities and Differences in Inner Speech," and "Changes in Inner Speech Following Random Sampling"). Finally, mean responses for happiness, anxiety, and temporal thinking were used to explore relations with generalized and randomly sampled inner speech. As this final analysis was exploratory, results were treated as significant at $p < 0.05$ (see "Relations to Mood and Temporal Thinking").

Results

Overall Characteristics of Inner Speech

All participants provided a full set of T1 data and responded to at least 25% of their app alerts (7/28 samples). The mean percentage of alerts responded to was 63.14% (SD = 17.10, Range = 29–93). The retest of general VISQ at T2 was also completed by 36 participants (see **Table 1** for mean scores on the VISQ).

Similarities and Differences in Inner Speech

Spearman correlations between generalized endorsements of inner speech and random sampling incident reports were significant for Condensed ($r = 0.69$, $df = 49$, $p < 0.001$) and Other People ($r = 0.46$, $df = 49$, $p < 0.001$), but only approached significance for Dialogic, given the use of an adjusted alpha value ($r = 0.30$, $df = 49$, $p = 0.031$). There was no correlation between generalized and random scores for Evaluative ($r = 0.03$, $df = 49$, $p = 0.851$).

When the total scores for inner speech were compared across generalized endorsements and randomly sampled reports, each of the subscales were significantly lower when an experience sampling method was used. As **Table 1** shows, the greatest discrepancies were for Dialogic (Wilcoxon's $Z = -4.88$, $df = 50$, $p < 0.001$) and Evaluative ($Z = -5.88$, $df = 51$, $p < 0.001$), followed by Condensed ($Z = -3.19$, $df = 48$, $p < 0.001$) and Other People ($Z = -2.82$, $df = 48$, $p = 0.005$).

When scores were compared between subscales, T1 reports of Evaluative inner speech were significantly higher than scores for Other People ($Z = -5.03$, $df = 48$, $p < 0.001$), Condensed ($Z = -3.54$, $df = 46$, $p < 0.001$), and, at trend level, Dialogic ($Z = -2.29$, $df = 47$, $p = 0.022$). This was not the case under random sampling, where the only pairwise comparison observed

to reach significance was between Dialogic and Other People (12.94 vs. 10.48, respectively; $Z = -2.52$, $df = 51$, $p = 0.012$). In addition, the average discrepancy between T1 and random-sampling was greatest for Evaluative scores compared to all other subscales (all $p < 0.002$).

Changes in Inner Speech Following Random Sampling

Test–retest reliability for T1 to T2 generalized endorsements of inner speech were significant and within an acceptable range for Dialogic ($r = 0.81$, $df = 34$, $p < 0.001$), Condensed ($r = 0.88$, $df = 34$, $p < 0.001$), and Other People ($r = 0.68$, $df = 34$, $p < 0.001$). Again, Evaluative was much less reliable, showing only a modest correlation between T1 and T2 ($r = 0.42$, $df = 34$, $p = 0.011$). There were no significant differences in overall VISQ scores between T1 and T2.

Relations to Mood and Temporal Thinking

No significant correlations were observed between mean levels of happiness and either generally endorsed or randomly sampled VISQ scores (all $r < 0.15$, $p > 0.30$). Greater anxiety scores were associated with Other People scores during random sampling ($r = 0.30$, $df = 49$, $p = 0.032$) and, at trend level, T1 endorsements for the same subscale ($r = 0.27$, $df = 49$, $p = 0.057$). Anxiety was also associated with endorsement of Evaluative inner speech at T1 ($r = 0.29$, $df = 49$, $p = 0.043$) but not during random sampling ($r = 0.05$, $df = 49$, $p = 0.716$). Temporal thinking (where higher scores indicated thinking about the future and lower scores indicated thinking about the past) was positively associated with Condensed inner speech ($r = 0.36$, $df = 49$, $p = 0.01$) and negatively associated with Evaluative inner speech ($r = -0.32$, $df = 49$, $p = 0.023$), both at T1. However, these relations were only observed during random sampling for Condensed inner speech ($r = 0.41$, $df = 49$, $p = 0.003$).

Discussion

The main finding of the present study was that generalized endorsements of inner speech characteristics in many cases did not reliably indicate what is reported via random sampling, contrary to our hypothesis. Generalized reports of inner speech characteristics, gathered by a validated questionnaire (the VISQ), appeared to elicit generally higher levels of endorsement than randomly sampled incident reports. Furthermore, and perhaps most importantly, the correlation between generally endorsed

TABLE 1 | Endorsement of inner speech characteristics at start of the study (T1), during random sampling, and at the end of the study (T2).

	T1 (n = 51)		Random sampling (n = 51)			T2 (n = 36)	
	M	**(SD, Range)**	**M**	**(SD, Range)**	**Mean SD**	**M**	**(SD, Range)**
Dialogic (Max 24)	17.53	(5.24, 6–24)	12.94	(4.52, 4–20.62)	1.44	16.50	(6.10, 4–24)
Evaluative (Max 24)	19.45	(3.78, 8–24)	12.06	(4.28, 4.36–21.45)	1.30	18.69	(3.96, 8–24)
Other People (Max 30)	13.75	(6.89, 5–26)	10.48	(5.99, 5–25.59)	0.96	14.28	(7.00, 5–28)
Condensed (Max 30)	15.20	(6.66, 5–29)	12.73	(6.02, 5–28.75)	0.99	14.08	(6.75, 5–27)

(i) Randomly sampled VISQ ratings could be non-integers due to averaging across samples; (ii) Mean SD reflects average standard deviation of within-subjects samples.

and randomly sampled inner speech depended on the specific kind of inner speech being measured: asking about evaluative and motivational inner speech, compared to other phenomenological characteristics, did not produce consistent self-reports at all between questionnaire and experience sampling methods of measurement.

Participants' general endorsements of dialog-like inner speech, other people in inner speech, and condensed or fragmentary inner speech showed good test–retest reliability between the start and end of the study, and at levels that were actually higher than in McCarthy-Jones and Fernyhough's (2011) original study. These three subscales also showed at least some correlation with randomly sampled levels of inner speech collected via experience sampling, but not at the levels of reliability seen for test–retest of the general questionnaire. At the same time, overall endorsement levels for varieties of inner speech were significantly lower during random sampling, supporting Hurlburt et al.'s (2013) argument that asking about inner speech in generalized terms may lead to inflated responses.

Why over-estimation would be occurring is an important question to answer. One thing to note is that the general VISQ does not actually ask about frequency of inner speech: instead, participants rate their level of agreement for what their inner speech is generally like, whenever it occurs. As Hurlburt et al. (2013) note, endorsement of items on this basis provides ambiguous information regarding whether a given experience of inner speech happens often, or happens infrequently but in such a way that makes a participant strongly identify with the scenario described. The random-sampling reports, in contrast, asked about characteristics of inner speech at the moment of the alert, even when inner speech may not have been occurring. Thus, to some degree, reports of inner speech phenomenology are bound to be lower whenever random sampling is used in this way.

However, even with frequency-based responding, overestimates of inner speech may still be expected from a generalized questionnaire. Self-report questionnaires are notoriously susceptible to various reporting biases, affecting both recall of the phenomenon in question and judgments about how often it occurs (e.g., Houtveen and Oei, 2007; Edmondson et al., 2013). The peak level of a behavior or experience, its level at the end of sampling, and its variability across time can all affect participant accuracy: for instance, participants with more variable chronic pain also tend to overestimate their average pain level compared to those with more consistent pain (Stone et al., 2005).

Consideration of such biases is important for interpreting the results regarding evaluative inner speech. While the other three subscales of the VISQ showed at least some evidence of reliability between the general questionnaire and random sampling, estimates of evaluative and motivational characteristics of inner speech were worryingly divergent. Generalized reports for this factor were not significantly related to random-sampling levels and showed relatively low test–retest reliability for the subsample who completed the VISQ again at the end of the study. This is in distinct contrast with McCarthy-Jones and Fernyhough's (2011) data, in which general scores for evaluative inner speech showed high test–retest reliability (0.80). Correspondingly, correlations

with mood and temporal thinking observed for this characteristic did not hold across generalized and randomly sampled measurements.

Here and in prior studies (McCarthy-Jones and Fernyhough, 2011; Alderson-Day et al., 2014), evaluative inner speech was the VISQ subscale that participants endorsed the most, and was the most likely to be over-estimated. Given prior evidence of variability effects (e.g., Stone et al., 2005), it could be that evaluative inner speech is harder to estimate because of its variance across time compared to other subscales. However, the mean SDs of each subscale do not support this idea (see **Table 1**): Dialogic inner speech was the most variable subscale, rather than Evaluative.

A second reporting bias could arise from the content of evaluative inner speech. Strongly valenced behaviors and states have often been reported to affect accuracy of recall (Shiffman et al., 2008). On both the VISQ (Alderson-Day et al., 2014) and other measures of inner speech (Brinthaupt et al., 2009), the tendency to engage in self-reflective and evaluative processes appears to be linked to negative beliefs and ideas about oneself. If so, evaluative inner speech, compared to other VISQ subscales, may have a greater salience to participants when they think about it in generalized terms, leading to its overestimation on questionnaires. Were this to be the case, discrepancies in inner speech reporting would be expected to be greatest for measures that specifically enquire about positive and negative statements in inner speech (such as the Self-Talk Inventory; Calvete et al., 2005), or in individuals with a tendency toward engaging in more negative, ruminative inner speech behaviors (such as people with depression; Nolen-Hoeksema, 2004).

Exploratory analysis of the relations between inner speech, mood, and temporal thinking indicated two avenues for future study. First, the presence of other people in inner speech—assessed via random sampling—was also associated with greater levels of momentary anxiety. This is consistent with this factor being a general marker for psychopathology, as it has previously been observed to relate to increased trait anxiety, hallucination-proneness, and dissociative tendencies (McCarthy-Jones and Fernyhough, 2011; Alderson-Day et al., 2014). This was also evident for generalized endorsement of other people in inner speech, but only at trend level, suggesting again that random sampling could provide a more accurate assessment of this particular characteristic. Second, higher levels of condensed inner speech (via generalized endorsement and randomly sampled reports) were associated with thinking about the future rather than the past. This is in line with links between inner speech and future thinking observed by D'Argembeau et al. (2011), but suggests that the kind of inner speech being used may differ depending on whether someone is thinking about the past, present, or the future. Condensed inner speech is proposed by Fernyhough (2004) to represent a syntactically and semantically abbreviated form of verbal thinking that results from the internalization of external speech. Its counterpart is expanded inner speech, in which internal talk involves full words and sentences. Speculatively, it is possible that condensed thinking could aid future planning, while more expanded inner speech could be involved in reflecting on the past and specific recall, including reconstructing past events in the greater detail afforded by an expanded linguistic code.

In contrast, no associations were observed between happiness ratings and VISQ scores. This is inconsistent with D'Argembeau et al.'s (2011) finding that inner speech diary reports were more likely to be associated with negative rather than positive thinking, suggesting again that generalized endorsements about inner speech do not show consistent evidence of validity. Happiness has previously been negatively associated with mind-wandering reports collected via experience sampling (Killingsworth and Gilbert, 2010), suggesting that attentional control factors may be more important factors for understanding mood variation than the presence or absence of inner speech.

Some limitations to this study must be acknowledged. First, only one measure of inner speech —the VISQ—was used here, and it is possible that the discrepancy between inner speech reports observed is an artifact of this specific scale rather than inner speech measures more generally. Given the limited validity of other major scales (Uttl et al., 2011), it is not clear that alternative measures would necessarily have performed better. However, the VISQ could definitely be improved: as noted above, the present version of the scale asks participants to answer based on their agreement with general statements rather than specifically indicating frequency. To fully test why incidence estimates of inner speech differ so much between, for example, the studies of Klinger and Cox (1987) and Heavey and Hurlburt (2008) requires an index of inner speech that also asks about the frequency of particular experiences. We are currently in the process of adapting and expanding the VISQ for this purpose, which should remove some of the ambiguity in participants' responses noted by Hurlburt et al. (2013), and may lead to more accurate generalized data.

Second, for clarity of examining relations between generalized and randomly sampled inner speech, responses to momentary samples were averaged in the present study to provide mean scores. This, however, undoubtedly obscures the level of complexity inherent in experience sampling data. A key question for inner speech research is how it may vary across time in relation to mood and activity factors; further analysis with a larger sample will allow this to be assessed. Third, as is the case for the large majority of recent studies on inner speech, the data collected here are limited to an undergraduate student sample and do not necessarily reflect inner speech characteristics in the general population. The use of research apps like Inner Life should allow researchers to go beyond university-based samples and obtain a more accurate picture of the heterogeneity of inner speech.

Finally, it is important to note that any self-report method, whether gathered by questionnaire or momentary assessment, may be affected by reporting bias. A method such as DES differs from standard self-report techniques and most other experience sampling methods in its attempt to bracket experimenters' presuppositions and iteratively train participants to avoid their own. Hurlburt et al. (2013) argue that any self-report method is likely to provide inaccurate data on inner experience unless it attempts to do something similar. Notwithstanding the importance of such considerations, we argue here and elsewhere (Alderson-Day and Fernyhough, 2014) that a combination of methods is needed to examine inner speech both on an in-depth, individual level (as in DES) and in larger, representative samples. Examining how data from self-reports may change in their reliability and validity with iterative training is also important for establishing whether people can 'improve' in their reporting of inner experience, as DES would hold, although in the present dataset, there were no clear signs of reactivity in response to use of random sampling (cf. Runyan et al., 2013).

In summary, the present article describes the first app-based study of everyday inner speech. Generalized estimates of inner speech can in some cases be reliable indicators of day-to-day characteristics of inner speech, but this varies considerably depending on the kind of self-talk being asked about. It seems likely that memory biases and other confounds affect self-reports about inner speech, particularly for its evaluative and motivational features. This does not mean that self-reports of inner speech are entirely inaccurate, but it strongly suggests that when we ask about inner speech, participants are reporting on the kinds of subjective experiences and processes that are salient and important to them.

Acknowledgments

References

Alderson-Day, B., and Fernyhough, C. (2014). More than one voice: investigating the phenomenological properties of inner speech requires a variety of methods. *Conscious. Cogn.* 24, 113–114. doi: 10.1016/j.concog.2013.12.012

Alderson-Day, B., McCarthy-Jones, S., Bedford, S., Collins, H., Dunne, H., Rooke, C., et al. (2014). Shot through with voices: dissociation mediates the relationship between varieties of inner speech and auditory hallucination proneness. *Conscious. Cogn.* 27, 288–296. doi: 10.1016/j.concog.2014.05.010

Baars, B. (2003). How brain reveals mind: neural studies support the fundamental role of conscious experience. *J. Conscious. Stud.* 10, 100–114.

Brinthaupt, T. M., Hein, M. B., and Kramer, T. E. (2009). The self-talk scale: development, factor analysis, and validation. *J. Pers. Assess.* 91, 82–92. doi: 10.1080/00223890802484498

Calvete, E., Estévez, A., Landín, C., Martínez, Y., Cardeñoso, O., Villardón, L., et al. (2005). Self-talk and affective problems in college students: valence of thinking and cognitive content specificity. *Span. J. Psychol.* 8, 56–67. doi: 10.1017/S1138741600004960

Csikszentmihalyi, M., and Larson, R. (1987). Validity and reliability of the Experience-Sampling method. *J. Nerv. Ment. Dis.* 175, 526–536. doi: 10.1097/00005053-198709000-00004

D'Argembeau, A., Renaud, O., and Van der Linden, M. (2011). Frequency, characteristics and functions of future-oriented thoughts in daily life. *Appl. Cogn. Psychol.* 25, 96–103. doi: 10.1002/acp.1647

Duncan, R. M., and Cheyne, J. A. (1999). Incidence and functions of self-reported private speech in young adults: a self-verbalization questionnaire. *Can. J. Behav. Sci. Rev.* 31, 133–136. doi: 10.1037/h0087081

Edmondson, D., Shaffer, J. A., Chaplin, W. F., Burg, M. M., Stone, A. A., and Schwartz, J. E. (2013). Trait anxiety and trait anger measured by ecological momentary assessment and their correspondence with traditional trait questionnaires. *J. Res. Pers.* 47, 843–852. doi: 10.1016/j.jrp.2013.08.005

Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas Psychol.* 22, 49–68. doi: 10.1016/j.newideapsych.2004.09.001

Goldstein, G., and Kenen, R. (1988). "Internal dialogue" in a "normal" population: the implications for health promotion. *Health Promot. Int.* 3, 249–257. doi: 10.1093/heapro/3.3.249

Heavey, C. L., and Hurlburt, R. T. (2008). The phenomena of inner experience. *Conscious. Cogn.* 17, 798–810. doi: 10.1016/j.concog.2007.12.006

Houtveen, J. H., and Oei, N. Y. L. (2007). Recall bias in reporting medically unexplained symptoms comes from semantic memory. *J. Psychol. Res.* 62, 277–282. doi: 10.1016/j.jpsychores.2006.11.006

Hurlburt, R. T., Heavey, C. L., and Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Conscious. Cogn.* 22, 1477–1494. doi: 10.1016/j.concog.2013.10.003

Killingsworth, M. A., and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science* 330, 932–932. doi: 10.1126/science.1192439

Klinger, E., and Cox, W. M. (1987). Dimensions of thought flow in everyday life. *Imagin. Cogn. Personal.* 7, 105–128. doi: 10.2190/7K24-G343-MTQW-115V

McCarthy-Jones, S., and Fernyhough, C. (2011). The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Conscious. Cogn.* 20, 1586–1593. doi: 10.1016/j.concog.2011.08.005

Morin, A. (2005). Possible links between self-awareness and inner speech: theoretical background, underlying mechanisms, and empirical evidence. *J. Conscious. Stud.* 12, 115–134.

Morin, A., Uttl, B., and Hamper, B. (2011). Self-reported frequency, content, and functions of inner speech. *Procedia Soc. Behav. Sci.* 30, 1714–1718. doi: 10.1016/j.sbspro.2011.10.331

Nolen-Hoeksema, S. (2004). "The response styles theory," in *Depressive Rumination: Nature, Theory and Treatment*, eds C. Papageorgiou and A. Wells (Chichester: Wiley), 107–124.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., and Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behav. Brain Res.* 261, 220–239. doi: 10.1016/j.bbr.2013.12.034

Runyan, J. D., Steenbergh, T. A., Bainbridge, C., Daugherty, D. A., Oke, L., and Fry, B. N. (2013). A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness. *PLoS ONE* 8:e71325. doi: 10.1371/journal.pone.0071325

Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4, 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Siegrist, M. (1995). Inner speech as a cognitive process mediating self-consciousness and inhibiting self-deception. *Psychol. Rep.* 76, 259–265. doi: 10.2466/pr0.1995.76.1.259

Sokolov, A. N. (1975). *Inner Speech and Thought*. New York, NY: Plenum Publishing Corporation. doi: 10.1007/978-1-4684-1701-2

Stone, A. A., Schwartz, J. E., Broderick, J. E., and Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Personal. Soc. Psychol. Bull.* 31, 1340–1346. doi: 10.1177/0146167205275615

Uttl, B., Morin, A., and Hamper, B. (2011). Are inner speech self-report questionnaires reliable and valid? *Procedia Soc. Behav. Sci.* 30, 1719–1723. doi: 10.1016/j.sbspro.2011.10.332

Vygotsky, L. S. (1987). *Thinking and Speech. The Collected Works of Lev Vygotsky*, Vol. 1. New York, NY: Plenum Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Psychology

Virtues, ecological momentary assessment/intervention and smartphone technology

Jason D. Runyan * and *Ellen G. Steinke*

Psychology Department, Indiana Wesleyan University, Marion, IN, USA

Virtues, broadly understood as stable and robust dispositions for certain responses across morally relevant situations, have been a growing topic of interest in psychology. A central topic of discussion has been whether studies showing that situations can strongly influence our responses provide evidence against the existence of virtues (as a kind of stable and robust disposition). In this review, we examine reasons for thinking that the prevailing methods for examining situational influences are limited in their ability to test dispositional stability and robustness; or, then, whether virtues exist. We make the case that these limitations can be addressed by aggregating repeated, cross-situational assessments of environmental, psychological and physiological variables within everyday life—a form of assessment often called ecological momentary assessment (EMA, or experience sampling). We, then, examine how advances in smartphone application (app) technology, and their mass adoption, make these mobile devices an unprecedented vehicle for EMA and, thus, the psychological study of virtue. We, additionally, examine how smartphones might be used for virtue development by promoting changes in thought and behavior within daily life; a technique often called ecological momentary intervention (EMI). While EMA/I have become widely employed since the 1980s for the purposes of understanding and promoting change amongst clinical populations, few EMA/I studies have been devoted to understanding or promoting virtues within non-clinical populations. Further, most EMA/I studies have relied on journaling, PDAs, phone calls and/or text messaging systems. We explore how smartphone app technology provides a means of making EMA a more robust psychological method, EMI a more robust way of promoting positive change, and, as a result, opens up new possibilities for studying and promoting virtues.

Keywords: experience sampling, mindfulness, self-awareness, habits, automaticity, character traits, virtues, dispositions

Introduction

Over the past 15 years, virtues have received increased attention in the psychological sciences. This has, in large part, been a result of the positive psychology movement (Seligman and Csikszentmihalyi, 2000; Seligman et al., 2005). Positive psychology is a subfield of psychological science devoted to a deliberate attentiveness to human flourishing and its promotion. And virtues have been a central focus of positive psychology (Seligman et al., 2005; Kristjánsson, 2013; Worthington et al., 2014).

This increased attention to virtues, and positive psychology in general, has not gone without criticism (e.g., Gable et al., 2005; Held, 2005; Sundararajan, 2005; Kristjánsson, 2010, 2013; McNulty and Fincham, 2012). In this review, our goal is not to defend positive psychology. We, however, propose that there is virtue in the psychological study of virtues. We, further, propose that smartphone technology opens up a new means of studying, and possibly promoting, virtue. In making our case, we discuss:

- philosophical and psychological work concerning virtue stability (persistence over time) and robustness (cross-situational persistence);
- smartphone application (app) advances pertaining to the measurement of human (and physiological) activity and experience within daily life; and
- the use of mobile app technology for ecological momentary assessment and intervention as it pertains to the study of virtues.

Virtues and Situational Studies: Questions of Stability and Robustness

Dating at least as far back as Aristotle's analytic treatment of virtues, there has been a long history of understanding virtues as a kind of disposition (*hexis*). This Aristotelian conception is often referred to as the traditional conception (cf. Timpe and Boyd, 2014); and, following a marked decline in interest, virtues understood along Aristotelian lines have received renewed attention in moral philosophy. This renewed attention is in large part a result of the influence of works by Anscombe (1958), Foot (1978), and MacIntyre (1984).

According to a traditional Aristotelian conception, virtues are human excellences understood as a *subclass* of psychological disposition, or that is, ability to behave or think in certain ways, or have certain psychological responses, across relevant situations (Aristotle, 2000 I.13; Aristotle *EE* II.1218-20). Expressing virtues is constitutive of a flourishing (*eudaimonic*) life, or, that is, a deeply fulfilling, well-lived life of growth (cf. Aristotle *EE* I.7-II.1; Fowers, 2012). While a person can be virtuous without flourishing, and even while suffering, a person cannot flourish without being virtuous[1]. In the way a good ax cuts wood well, a flourishing person expresses virtues (Aristotle *EE* II.1213-34). And virtues are expressed by actively doing or thinking certain things, or by feeling certain emotions or having certain passive responses (e.g., refraining from certain behavior), in morally relevant situations (cf. Jayawickreme and Chemero, 2008).

Virtues, understood as a subclass of disposition, have been recognized as having three important characteristics (Timpe and Boyd, 2014). First, they are relatively *stable*. Virtues tend to

persist over some period of time. Thus, we generally cannot be sure whether someone has a virtue until they have expressed it on multiple occasions. Similarly, we generally cannot be sure whether someone has a virtue until they have expressed it under various virtue-relevant situations. And this relates to the second characteristic of virtues. Virtues are relatively *robust* in the sense that they are consistently expressed across a range of situations. Third, virtues are *interconnected* in the sense that having one virtue increases the probability of having others (cf. Watson, 1984; MacIntyre, 1999; Annas, 2011).

In modern psychology, Allport provided an early treatment that supported the conception of virtues as a kind of relatively stable, robust and interconnected disposition or trait (cf. Allport, 1960)[2]. However, more recently, a number of moral psychologists have argued that there are good experimental grounds for thinking virtues do not actually exist (cf. Nahmias, 2010). The argument is that virtues are no more than unmaterialized ideals. The grounds for this argument come from studies indicating that, to a significant degree, a person's situation can influence their conduct and thought without them being aware of it (Hunt, 1965; Mischel, 1968; Ross and Nisbett, 1991; Doris, 1998, 2002; Harman, 1999). In one of the most well-known of these studies—Zimbardo's Stanford Prison Experiment—college students began exhibiting guard-like or inmate-like behavior after only a few days of taking on the role of either a guard or a prisoner in a mock prison (Haney et al., 1973). In another well-known experiment, Milgram (1963) found that a majority of participants would administer what they thought to be a potentially lethal shock to individuals they had never met if ordered to by an experimenter as part of what was presented to them as a scientific study (see also Hartshorne and May, 1928; Asch, 1951; Isen and Levin, 1972). Additionally, over the past 20 years, a wealth of studies have indicated that priming individuals by having them, for example, read words with either prosocial or antisocial connotations, or handle certain kinds of objects (e.g., smooth, rough, heavy, hot, cold), can influence their subsequent behavior and judgments without them realizing it (e.g., Williams and Bargh, 2008; Ackerman et al., 2010; Bargh and Shalev, 2012).

---

[1]It should be noted that the Aristotelian tradition is a eudaimonic tradition in which happiness, or flourishing, is analyzed in terms of living a fulfilled and deeply satisfying life of growth through expressing virtues (Kenny, 2011, pp. xii–xiv). This tradition is in opposition to hedonic traditions in which happiness is analyzed in terms of affect or subjective well-being (cf. Deci and Ryan, 2008a). At the same time, accepting a eudaimonic analysis does not entail, or typically involve, thinking it can never be the case that a person who is exercising virtues is not flourishing; or that a flourishing life will not include subjective well-being or pleasure.

[2]Two points should be made here: (1) First, Allport, and many psychologists since, talk of traits (i.e., personality traits, character traits) rather than psychological dispositions; and, while, "trait" and "disposition" are commonly used interchangeably, traits are often thought of as a type of more stable disposition. Having said this, there are variations in the use of the term "disposition." Sometimes, "disposition" and "trait" are both used to refer to any relatively stable and robust psychological characteristic, including deficiencies or the lack of certain psychological abilities (e.g., Eysenck and Eysenck, 1969; Watson and Clark, 1984; McCrae and Costa, 2003). Other times, often in philosophical psychology, a psychological disposition is understood as any psychological ability, some of which are not, in and of themselves, typically understood as a trait or aspect thereof (e.g., Annas, 2011; Anjum et al., 2013; Hyman, 2014). For our purposes here, it is not crucial to analyze various conceptions of psychological traits or dispositions, or how these conceptions overlap. It suffices to point out that, often, there is overlap in the use of "trait" and "disposition," and, for our purposes here, a psychological disposition is a psychological ability. A virtue is, then, a subclass of psychological disposition or ability as outlined above (also see Mumford, 1998). (2) Second, according to Allport (1960), the study of virtues involves value judgments and, thus, is a study for moral philosophy rather than psychological science. As will become clear, we take a different view.

The kinds of situational studies mentioned above have been thought to provide evidence against the existence of virtues as relatively stable and robust dispositions; that is, as dispositions consistently expressed across relevant situations over a period of time (Harman, 1999; Doris, 2002). However, while drawing attention to the extent and ways situations can influence individuals, to think these studies provide evidence against the existence of virtues, conceived of as a type of stable and robust disposition, is to infer too much. As Croom (2014) and others (e.g., Alzola, 2008) have recently pointed out, thinking these studies provide evidence against the existence of virtues results from: oversimplifying an Aristotelian conception of virtues as perfectly stable and robust dispositions; and/or drawing unwarranted conclusions from these studies.

First, as Anscombe (1958, p. 14) has pointed out, regardless of the virtues possessed by the average person, there may be a complete set of virtues each of which is possessed by some people. Virtues may be rare without being nonexistent; and, granted they exist, they are likely rare. Thus, when studying whether virtues exist, it is necessary to distinguish candidates for having a certain stable and robust disposition from other individuals in order to examine whether relevant situational influences have similar effects on both groups. So it is necessary to examine exemplars (e.g., Colby and Damon, 1992, 1999; Dunlop and Walker, 2013). Situational factors may not influence exemplars in the way they do the average person. Exemplars may express a virtue despite situational influences that make it difficult to do so. This, however, has not been tested in the situational studies purported to call the existence of virtues in question. As a result, the situational effects observed in these studies should not be generalized to the entire human population. Even if such situational effects are found in the majority, this does not provide evidence that virtues do not exist.

Second—and in keeping with the above—the situational studies purported to provide evidence that virtues are nonexistent do not even provide evidence that virtues are not possessed by a subgroup within the samples studied (cf. Miller, 2013). Thus, at most, these studies indicate that certain virtues are rare. For instance, in Darley and Batson's (1973) "Good Samaritan" study, only 10% of seminarians that participated in the study helped someone who appeared to need medical attention as they rushed to give a talk for which they were very late. While these observations might be taken to indicate that 90% of the seminarians lack a certain virtue, it clearly should not be taken to indicate that all of them do.

Third, many of the studies thought to call into question the existence of virtues have used undergraduate students who may still be developing in ways relevant to the development of virtues. It has been observed that the prefrontal cortex, and its connectivity to other regions, typically continues to develop up to late- or post-adolescence. This development is inversely correlated with novelty seeking (e.g., Pfefferbaum et al., 1994; Reiss and Havercamp, 1996; Sowell et al., 1999, 2003; Gogtay et al., 2004; Kelley et al., 2004; Segalowitz and Davies, 2004; Somerville et al., 2010, 2011) and impulsivity (Shannon et al., 2011), which are, in turn, likely to be inversely correlated with possessing *stable* and *robust*

psychological dispositions, including virtues. Thus, many of these studies are performed using samples that tend toward not having yet developed certain, stable psychological dispositions.

Fourth, to provide evidence that virtues do not exist, it would need to be shown that engaging in practices thought to contribute to virtue development does not mitigate situational influences on an individual's responses. However, none of the situational studies thought to bring the existence of virtues in question tests this.

Fifth, a substantial amount of evidence indicates that cross-situational consistencies in responses characteristic of relatively stable and robust dispositions—such as virtues—are often missed without the aggregation of repeated, cross-situational measurements (Dlugokinski and Firestone, 1973, 1974; Staub, 1974; Epstein, 1979, 1983; Rushton, 1980, 1984; Rushton et al., 1981, 1983; Fleeson, 2001; Furr, 2009). The reason is there are multiple responses characteristic of these kinds of dispositions and there is some degree of variability in their expression as a result of interfering factors (Fleeson and Noftle, 2008; Miller, 2013). Thus, using one or two situational tests to examine whether an individual possesses a disposition, or virtue, is an insufficient and unreliable approach to testing dispositional stability and robustness.

In sum, situational studies thought to provide evidence against the existence of virtues are not optimally designed to test whether virtues exist, and, thus, should not be taken to indicate they do not. There is, on the other hand, accumulating evidence that at least some individuals do possess virtues.

First, studies that aggregate repeated, cross-situational measurements of responses provide evidence that at least some individuals express relatively stable and robust psychological dispositions (Ozer, 1986; Fleeson, 2001, 2004; McNiel and Fleeson, 2006; John et al., 2008; Donnellan and Lucas, 2009), even if they are not impervious to situational influences (Fleeson, 2007; Fleeson and Noftle, 2008; Bleidorn, 2009). It is likely that some of these dispositions contribute to psychological health and growth and so, in keeping with a broadly conceived Aristotelian tradition, are virtues.

Second, it has been observed that at least some psychological states tend to fluctuate around a "set point" (e.g., Suh et al., 1996; Diener and Lucas, 1999; Diener et al., 2006; Keltner, 2009). Since a psychological state can be the expression of a disposition, this observation provides further evidence that relatively stable and robust psychological dispositions exist. And, again, it is likely that some of these dispositions contribute to psychological health and growth; and so—again in keeping with an Aristotelian tradition—are virtues.

Third, virtues can be possessed in degrees and the degree to which a virtue is possessed can be used to predict patterns of responses, such as the frequency of virtue-relevant responses across relevant situations (cf. Miller, 2013). At the same time, as with all dispositions, there are factors that can interfere with the expression of virtues. Thus, some variability in virtue expression across relevant situations is to be expected, and this variability can be used to measure degree of virtue possession (Miller, 2013).

Finally, as we will see in what follows, there is growing psychological and neurophysiological evidence suggesting that training can promote the development of relatively stable and robust psychological dispositions which contribute to flourishing; or, that is, to virtue development understood along Aristotelian lines.

In this review we make the case that advances in smartphone technology open up a new approach to the psychological study of relatively stable and robust dispositions. We also make the case that putting this technology to use in this capacity promises to add to mounting psychological and neurophysiological evidence that certain individuals possess virtues understood along Aristotelian lines. Having stated this, we should be careful not to undervalue longstanding, everyday evidence that, throughout history, certain individuals have consistently expressed virtues in spite of strong situational influences to the contrary (see Colby and Damon, 1992). For instance, steadfast individuals who hid Jews in Nazi Germany, or undermined the Nazi regime in other ways, have been well documented and readily come to mind (e.g., see Marsh, 2014). We should, instead, seek to learn from such exemplars. Further, we should be careful not to undervalue the fact virtues have remained a useful, relevant and commonplace construct for thousands of years.

Keeping the above in mind, we propose the integration of something ancient and something cutting-edge: the study of virtues and the use of smartphone app technology. We propose that recent advancements in smartphone app technology, and the mass adoption of this technology, opens up a new means of examining and developing virtues through ecological momentary assessment (EMA) and ecological momentary intervention (EMI), respectively. In the remainder of this review, we, first, introduce EMA and discuss how smartphone technology provides a vehicle for making EMA a robust psychological method. We, second, examine how smartphone EMA studies promise to add to our knowledge of virtues; and, in particular, virtue stability and robustness. We, then, introduce EMI and examine how smartphone technology provides a vehicle for making EMI a widespread and effective way of promoting positive change. Finally, we discuss the promise smartphone EMI holds as an effective means of promoting virtue development.

EMA and Smartphone Apps

Quantitative psychological studies have, traditionally, relied heavily on surveys and laboratory experiments. Both approaches have well-known and long-endured limitations. Surveys require people to make retrospective and often generalized judgments, which tend to be affected by memory limitations and recall biases (cf. Schwarz, 2007). Laboratory experiments do not occur within the context of a person's daily life; and context can influence a person's states and responses (cf. Hammond et al., 1998; Wilhelm et al., 2011), which raises questions about ecological validity (Shiffman et al., 2008). Additionally, since neither surveys nor laboratory experiments involve repeated, cross-situational sampling, both approaches fail to detect intrasubject variability within the context of an individual's everyday life (cf. Hammaker, 2012). This is particularly relevant with respect to

studying virtues since, as observed in the previous section, the extent to which an individual's states and responses vary across situations—as well as individual differences in this regard—are crucial to measure when examining virtue possession.

Though earlier analogs existed, ecological momentary assessment (EMA, also referred to as experience sampling or ambulatory assessment) developed in the 1980s as a way of addressing the limitations of traditional quantitative methods in psychological science (Csikszentmihalyi and Larson, 1987; Stone and Shiffman, 1994; Shiffman et al., 2008). In particular, it was developed as a form of assessment that allowed repeated sampling within the various situations of daily life. In EMA, individuals are prompted, at fixed or random times, to respond to questions about what they are presently doing and/or experiencing (or what they have done and/or experienced in the recent past), repeatedly, throughout a period of time within the course of their daily affairs. EMA has been implemented in a number of ways, including through the use of stopwatches and paper-and-pencil diaries, PDAs (personal digital assistants; e.g., PalmPilots), phone calls and text messages. However, progress in smartphone technology has, recently, opened up a new mode of EMA.

With the release of the iPhone OS2 operating system in 2008, smartphones that could run third-party applications, or "apps," began being used in the daily activities of millions of people. By 2009, with the release of the iPhone OS3 operating system, millions began carrying devices that could run multiple apps continuously in the background; and some of these apps could run without an internet connection. Today, a number of companies make smartphones with this capability and approximately 1.91 billion people carry these devices (eMarketer, 2015). It is projected that by 2018 this number will increase to 2.56 billion.

In more economically "developed" countries, the near ubiquitous use of smartphone apps makes EMA practical for widespread use. For the first time, EMA can be conducted in a robust and dynamic way by using a tool that is already a part of daily life for a large percentage of the population (Raento et al., 2009). The widespread use of smartphones opens up a means of collecting psychological data within the moments of daily life along with data collected through various types of sensors (e.g., global positioning systems (GPS), microphones, cameras, activity monitors, heart rate monitors). And, unlike with other modes of EMA, participants need not be trained to use a new device. Additionally, whereas using PDAs for EMA requires a certain amount of programming expertise, flexible smartphone app-based EMA systems are being developed and distributed that allow researchers to create their own EMA designs through a user-friendly web interface that requires no programming expertise (for a list see Conner, 2014; Konrath, 2015; also see **Table 1**)[3]. Further, smartphone app-based EMA systems have been created that automatically enter data into datasets as the data streams in from participants' smartphones.

---

[3] Most smartphone app-based EMA systems work on either Apple or Android devices, which make up the majority of smartphones in use. Only two work with both: illumivu and LifeData.

TABLE 1 | EMA peer-reviewed psychological studies using smartphone apps.

References	Assessment target	App	Sensors	Tailoring
Burns et al., 2011[a]	Depressive symptoms	Custom built	x	x
Palmier-Claus et al., 2012	Psychotic symptoms	Custom built		
Bossman et al., 2013	Affect and physical activity	Movisens	x	
von Haaren et al., 2013	Affect and physical activity	Movisens	x	
Walter et al., 2013	Mood and physical activity	Movisens		
Runyan et al., 2013	Time-management	iHabit		
Kirk et al., 2013	Illicit drug use	eMOCHA		
MacKerron and Mourato, 2013	Happiness	Mappiness		
Dunton et al., 2013[b]	Affect and physical activity amongst children	Custom built	x	
Watkins et al., 2014	Urges to smoke and location during attempts to quit smoking	Custom built	x	
Khor et al., 2014	Coping, behavior and emotion problems amongst adolescents with high-functioning autism	Mobiletype		
Garcia-Palacios et al., 2014	Chronic pain in fibromyalgia	Custom built		
Adams et al., 2014	Stress	SESAME		
Randall et al., 2014	Emotion, regulation strategies and music listening	MuPsych		
Gonzalez and Dulin, 2015	Alcohol use disorders	LBMI-A		
Ottaviani et al., 2015	Mind-wandering and cognitive rigidity	SurveyPocket and KoBo	x	
Bleidorn and Denissen, 2015	Virtues	Movisens		

These studies were located by searching Pubmed.com, and publications from peer-reviewed psychology journals listed in Scholar.google.com, up to February 27th, 2015 (search terms: "ecological momentary assessment" and "smartphone").
[a]In the Burns et al. (2011) study, depressive symptoms were assessed in order to predict mood for the purposes of tailoring when coping strategies were delivered. Thus, this was an EMA/I study.
[b]Dunton et al. (2013) used custom software downloaded onto a mobile phone rather than a standard smartphone app.

This increases the practicality of handling the relatively large amounts of data collected in EMA studies, which can easily approach ten thousand items of data.

In addition to making EMA more practical for widespread use, smartphone app-based EMA systems have been designed to ensure that participants respond to questions "in the moment," or, that is, soon after being alerted to do so. Some systems time stamp responses so that the time lapse between when a person is notified to answer a question and when they answer it can be calculated. Some systems also allow researchers to give participants a limited time window to respond to questions after being notified. These features help ensure that individuals are not merely giving a convenience sampling, which has been an issue with other modes of EMA (e.g., Stone et al., 2002).

As shown in **Table 1**, since 2011, there have been a number of psychological EMA studies conducted using smartphone apps. Several of these involved acquiring data through environmental, activity and/or physiological sensors together with self-report. One has involved tailoring assessments to the individual, and their moment-to-moment experiences, in response to their location and activity level.

To date, most app-based EMA studies have been conducted on clinical populations and many have been pilot studies. There have, however, been a number of non app-based EMA studies that have examined momentary dispositional expressions (see **Table 2**). To our knowledge, only one EMA study (published after this manuscript was under review) has systematically focused

on assessing virtues using momentary responses (see Bleidorn and Denissen, 2015). Nevertheless, EMA provides a means of repeatedly measuring an individual's states, experiences and responses, as well as the extent to which these vary, using multiple measures throughout the moments and situations of everyday life. And, as we saw in the previous section, repeated, cross-situational sampling using an aggregate of measures is crucial for the psychological study of virtues; and, in particular, for testing dispositional stability and robustness. Further, as Wichers (2014) has recently observed, measuring moment-to-moment states and events can provide insight concerning patterns contributing to the development of enduring unhealthy or healthy mental conditions. Thus, app-based EMA—as a means of measuring moment-to-moment states and events—provides an unprecedented vehicle for studying dispositions, including those that contribute to psychological well-being; i.e., virtues.

Virtues and App-Based EMA

Traditionally, in Western thought, *wisdom*, *justice*, *temperance*, and *courage* have been thought of as "cardinal" (derived from the Latin "cardo" meaning hinge), or principal, virtues (e.g., Wisdom of Solomon 8:7; Plato, 380 BC/1991; Ambrose et al., (377 AD/1961)). Aristotle, however, influentially extended this list and understood virtues to be optimal points between deficiencies and excesses (cf. Aristotle *EE*; see **Table 3**). Additionally, within Christian theology, *faith*, *hope,* and *love* (or *charity*) have

TABLE 2 | EMA peer-reviewed psychological studies targeting momentary dispositional expression within non-clinical populations.

Study	Target	EMA mode
Marco and Suls, 1993	Neuroticism	Paper and pencil
Stone et al., 1998; Schwartz et al., 1999; Roesch et al., 2010	Coping	PDA, PDA, Internet daily diary, respectively
Räikkönen et al., 1999; Vella et al., 2012; Demarble et al., 2014	Hostility	PDA and Monitoring devices for ECG and blood pressure
D'Antono et al., 2001	Agreeableness	Paper and pencil
Brown and Ryan, 2003; Levesque and Brown, 2007	Mindfulness	Paper and pencil
Conner and Barrett, 2005	Implicit self-attitudes and negative feeling states	PDA
Kane et al., 2007	Working memory capacity and mind-wandering	PDA
Moberly and Watkins, 2008; Huffziger et al., 2013	Ruminative self-focus	Paper and pencil, PDA, respectively
Conway et al., 2009	Empathy, altruism, helping behavior and affect	PDA
Burt and Donnellan, 2010	Antisocial behavior and acting-out	PDA
Minbashian et al., 2010	Conscientiousness	PDA
Fay and Sonnentag, 2012	Trait affect and proactive behavior	PDA
Bruehl et al., 2012	Anger and chronic pain intensity	PDA
Schwerdtfeger and Scheel, 2012	Self-esteem and cardiac vagal tone	PDA; Monitoring devices for ECG and bodily movement
Hofmann et al., 2012; Lopez et al., 2014	Self-control	PDA
Edmondson et al., 2013	Anxiety and anger	PDA
Silvia et al., 2014	Creativity	Cell phone delivered surveys
aan het Rot et al., 2015	Impulsivity, quarrelsomeness and agreeableness	Paper and pencil

These studies were located by searching Pubmed.com, and publications from peer-reviewed journals listed in Scholar.google.com, up to February 27th, 2015 (search terms: "ecological momentary assessment" and "virtue," or "trait," or "disposition"). Only studies targeting momentary dispositional expressions in non-clinical populations are included.

TABLE 3 | Aristotle's list of virtues.

Virtue (Mean)	Deficiency	Excess
Courage	Cowardice	Foolhardiness
Temperance	Insensibility	Intemperance
Liberality (Generosity in small matters)	Illiberality	Prodigality
Magnificence (Generosity in large matters)	Shabbiness	Extravagance
Self-worth	Diffidence	Vanity
Dignity	Servility	Churlishness
Gentleness	Impassivity	Irascibility
Candor	Dissembling	Boastfulness
Justice	Loss	Gain
Friendliness	Surliness	Flattery
Modesty	Shamelessness	Shyness
Righteous indignation	Malicious enjoyment	Envy
Wisdom	Naivety	Cunning
Hardiness	Softness	Toughness

Aristotle understood virtues to be means between the vices of excess and deficiency. This table lists Aristotle's virtues along with their corresponding excess and deficiency (adapted from Kenny's (2011) translation of Eudemian Ethics; note: Aristotle develops a slightly different list in the Nicomachean Ethics).

TABLE 4 | Peterson and Seligman's (2004) "Virtues in Action" classification of virtues.

Wisdom	Creativity, Curiosity, Open-mindedness, Love of learning, Perspective
Courage	Authenticity, Bravery, Persistence, Zest
Humanity	Kindness, Love, Social intelligence
Justice	Fairness, Leadership, Teamwork
Temperance	Forgiveness, Modesty, Prudence, Self-regulation
Transcendence	Gratitude, Hope, Humor, Religiousness

This table has been adapted from Seligman et al. (2005), which contains a description of the various virtue subtypes.

courage, humanity, justice, temperance, and *transcendence* (see **Table 4**). And there has been some indication that rankings of these characteristics strongly correlate across many countries ($n = 54$) and, to some extent, transcend ethnic, cultural and religious differences (Park et al., 2006; but see van Oudenhoven et al., 2012).

However, rather than understanding virtues as a prescribed set of characteristics, following the broadly conceived Aristotelian conception we outlined earlier, we understand virtues to be a kind of relatively stable and robust psychological disposition the expression of which contributes to a fulfilling, well-lived life of growth; or, that is, to a flourishing life. Whatever else a fully flourishing life may involve, such a life involves psychological growth, psychological (*eudaimonic*) well-being and physical health (cf. Ryan and Deci, 2001; Keyes, 2007; Ryff

traditionally been upheld as key "theological virtues" (e.g., 1 Corinthians 13:13; Aquinas, 1274/1948).

Recently, the case has been made that six "overarching" characteristics are widely upheld as virtues across most cultures (Peterson and Seligman, 2004; Dahlsgaard et al., 2005; Seligman et al., 2005; but see Shryack et al., 2010). These are: *wisdom,*

and Singer, 2008; Deci and Ryan, 2008a)[4]. In this case, since psychological growth, psychological well-being and physical health are measurable, which dispositions contribute to a flourishing life, and, thus, should be included in a list of virtues, can be empirically studied. What should be considered a virtue is also an important matter since a flourishing life is obviously desirable. That being said, it should be kept in mind that dispositions which contribute to a flourishing life may lead to flourishing under a certain range of circumstances without leading to flourishing under all circumstances. For instance, certain characteristics may contribute to flourishing *only* when possessed by a critical number of individuals within a social group. Further, characteristics may contribute to flourishing when possessed in clusters but not on their own. Thus, under certain circumstances, an individual may suffer despite, and even as a result of, expressing virtue. Likewise, an individual may not experience physical health despite expressing virtue.

Over the past several years, flourishing has received increased attention in psychological science, and a number of studies have identified psychological characteristics that correlate with psychological growth, psychological well-being and/or physical health. Dispositional resilience has been positively correlated with successful adaptation to life stress (Ong et al., 2006) and with the well-being of widows (O'Rourke, 2004; Rossi et al., 2007). Additionally, dispositional mindfulness has been positively associated with both psychological well-being and physical health (e.g., Bernstein et al., 2011; Baer et al., 2012; Bowlin and Baer, 2012; Tamagawa et al., 2013); and individual differences in dispositional mindfulness predict psychological health (e.g., Baer, 2003; Baer et al., 2004). Other dispositions positively associated with psychological and/or physical health, include gratitude (e.g., Wood et al., 2010; Emmons and Mishra, 2011), optimism (e.g., Scheier and Carver, 1987; Scheier et al., 1989, 2001; Engberg et al., 2013; Carver and Scheier, 2014; He et al., 2014), self-efficacy (e.g., Bandura, 2004; Luszczynska et al., 2005), compassion (e.g., MacBeth and Gumley, 2012), altruism (e.g., Brown et al., 2009), self-regulation (e.g., Nix et al., 1999; Wrosch et al., 2003; Deci and Ryan, 2008b; Simon and Durand-Bush, 2014), forgiving (e.g., Berry and Worthington, 2001; Farrow et al., 2001; Maltby and Day, 2001; Seybold and Hill, 2001; Lawler-Row, 2010), spirituality (e.g., Hill and Pargament, 2003; Miller and Thoresen, 2003; Kuo et al., 2014; Reutter and Bigatti, 2014), religiosity (e.g., Hummer et al., 1999; McCullough et al., 2000; Oman and Thoresen, 2005; Greenfield and Marks, 2007; Park, 2007), and wisdom (e.g., Webster and Deng, 2014).

Most of the studies associating dispositions with flourishing—including those mentioned above—rely on surveys to measure the disposition in question. They examine associations between the possession of these dispositions and some measure, or correlate, of flourishing. This approach, however, is not ideal for measuring, or then studying, dispositions for several reasons.

First, surveys assessing dispositions do not involve measuring the *expression* of dispositions within the context of an individual's daily life; or allow directly associating this expression with

flourishing. Rather, as we noted at the beginning of the previous section, surveys ask for generalized retrospective judgments removed from a person's daily context, which are susceptible to recall biases.

Second, as they do not involve repeated, cross-situational sampling, surveys assessing dispositions cannot effectively measure *intrasubject variability* in the expression of a disposition. As a result, this approach does not provide an effective means of measuring dispositional stability or robustness, which, as we have already seen, is paramount to the psychology study of virtues.

In contrast, EMA allows: (1) the detection of dispositional expression, and its correlates, within the context of daily life using multiple measures; and (2) the measurement of dispositional stability and robustness through repeated cross-situational sampling. EMA, thus, provides (3) a more thorough and direct means of examining the relationship between dispositions and flourishing than traditional approaches that rely on surveys. We will discuss points (1)–(3) in sequential order.

(1) Through the incorporation of environmental, activity, and physiological sensors, app-based EMA opens up various ways of detecting:

(i)   the expression of virtues;
(ii) correlates of having or expressing these dispositions; and
(iii) the relevant features of the situations in which they are expressed.

Virtues can be expressed by psychological states (e.g., emotional states, motivational states) as well as by what people think and do (e.g., Bartlett and DeSteno, 2006; DeSteno et al., 2010). Thus, EMA can be used to detect virtue expression by asking people questions pertaining to their recent or current psychological states, experiences, thought life and/or behavior (for e.g., see studies listed in **Table 2**) as well as by administering brief psychological tests (cf. Schlicht et al., 2013). An EMA app can prompt individuals to respond to questions, or take brief tests, repeatedly at various moments, and across various situations, throughout the day. As a result, rather than asking people to make generalized retrospective judgments (e.g., "In the past month, I have…"), an individual can be asked about their current or recent states, experiences or conduct (e.g., "Over the past hour, I have…"). And survey-style instruments assessing virtues might be adapted so that, rather than asking for generalized judgments, they ask for reports concerning the present or recent past (cf. Fleeson, 2001).

For example, Hofmann et al. (2014) recently used EMA to repeatedly prompt people at random times over a 3-day period to report moral and immoral behavior over the previous hour. This allowed for the detection of patterns in moral behavior (e.g., social contagion, moral licensing) and awareness (e.g., a relative tendency to note others' immoral rather than moral behavior). More recently, Bleidorn and Denissen (2015) took adjectives associated with, and listed in, Peterson and Seligman's (2004) six *Virtues in Action* classifications (see **Table 4**) that could be meaningfully inserted into the following sentence: "I behaved particularly… during the last hour." They, then, used app-based EMA to deliver these sentences to participants up to six times a

---

[4]Psychological well-being is not to be confused with subjective well-being (for e.g. see Ryff and Singer, 2008; Ryff, 2013). See note 1.

day over a 10-day period in order to have them rate their behavior in the past hour. Amongst their participants—working mothers and fathers—they found that an individual's average virtue rating, the degree of variability in their rating, and the way an individual typically responded in certain contexts were relatively stable.

In addition to delivering questions, smartphones can be used to randomly capture conversations or other activities (e.g., Mehl et al., 2012). Further, prompts can be given immediately after these recordings asking individuals to report their states, experiences or thoughts; and/or asking them to upload a picture or video of their surroundings. This allows the association of life events and situations with momentary states, experiences and/or responses that would, otherwise, be forgotten once individuals are more temporally and spatially removed from the event or situation. It also allows the capture of contextual details that an individual would, otherwise, be unaware of or forget.

Activity sensors (e.g., Fitbit, Polo tech, Apple Watch; Moviesens) can also be used to record any physical and physiological activities correlated with an individual's momentary responses, states and/or experiences (e.g., D'Antono et al., 2001; Schwerdtfeger and Scheel, 2012; Bossman et al., 2013; von Haaren et al., 2013; Demarble et al., 2014; Dunton et al., 2014). Participants might, further, be prompted to take saliva samples in order to examine potential biochemical correlates (e.g., cortisol, oxytocin or progesterone levels) of dispositional expression (e.g., Brown et al., 2008; Entringer et al., 2011; Koven and Max, 2014). Or they might take a pharmacological agent (e.g., tryptophan, an anti-anxiolytic, an anti-depressant) while participating in an EMA study targeting dispositional expression (cf. Moskowitz et al., 2001; Moskowitz and Young, 2006).

In the near future, app-based EMA will also allow the isolation of neurophysiological correlates of having and/or expressing certain dispositions within certain situations. For example, mobile electroencephalography (EEG) caps (e.g., Eegosports) could be synced with an EMA app in order to record event-related potentials, or preparatory neural activities such as readiness-potentials (RPs); i.e., relative changes in the activity of the primary motor cortex and surrounding regions associated with preparedness to act (e.g., Freude and Ullsperger, 1987; Coles et al., 1988; Coles, 1989; Deecke et al., 1990; Shibasaki and Hallett, 2006; Ibanez et al., 2012; Nachev and Hacker, 2014). Through these means we might find that, when an individual has a certain disposition, certain preparatory activities occur under certain situations. Similarly, app-based EMA that incorporates mobile EEG might allow the detection of other neurophysiological correlates of expressing dispositions similar to those recently measured for forgiving using *f* MRI, where an increase in activity in the angular gyrus was associated with forgiving (see **Figure 1**).

(2) By allowing various ways of detecting (i)–(iii)—listed above—smartphone app-based EMA provides a vehicle for more direct and repeated measurement of disposition-relevant responses across various daily situations using an aggregate of measures. It, thus, provides a means of collecting cross-situational data to populate a frequency distribution of an individual's disposition-relevant responses organized by the degree to which each expresses the disposition in question. From
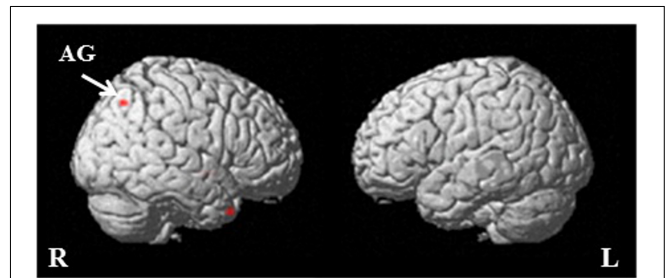
FIGURE 1 | Neural differences between forgiving and not forgiving as measured by fMRI. The red shows where increased activity was observed when participants forgave in comparison to when they did not (taken from Strang et al., 2014). AG, angular gyrus; R, right; L, left.

this distribution, a mean score for an individual's dispositional expression and the variability of this expression can be calculated (Fleeson and Noftle, 2008). In this way, EMA provides a way of measuring the typical degree to which, and consistency with which, an individual expresses a disposition throughout the relevant situations of their daily life over a period of time. So it provides a means of directly measuring the *stability* and *robustness* of a disposition, or virtue. In so doing, it provides a way for assessing not simply *whether* an individual has a virtue but the *degree* to which they have a virtue.

We should expect individuals who possess a certain virtue to *typically* express that virtue across a certain range of situations (Jayawickreme and Chemero, 2008). That is, given an Aristotelian conception, we should expect a virtue to be, to a certain degree, stable and robust. However, similar to the way the expression of other relatively stable and robust dispositions have been observed to vary (Ozer, 1986; McNiel and Fleeson, 2006; John et al., 2008; Donnellan and Lucas, 2009), some variability in the expression of a virtue should also be expected (see Miller, 2013). Nevertheless, the stronger, or more formed, a virtue, the more consistency there will be in its expression across relevant situations. This is because the stronger a virtue, the more frequently it is expressed in demanding situations, and despite interfering factors (Miller, 2013). So, after repeated cross-situational sampling of virtue-relevant responses, the degree to which an individual has a virtue can be measured as a function of the individual's mean score for its expression and the variability with which they express the virtue across relevant situations (cf. Fleeson, 2001; see **Figure 2**).

Thus, to recapitulate, EMA provides a way of repeatedly measuring virtue-relevant responses across various situations. This allows the measurement of the degree to which a person has a virtue along two dimensions: the degree to which they typically express the virtue and the consistency with which they express the virtue. These observations are in keeping with observations made concerning distributions of responses relevant to the expression of dispositions more generally (e.g., Mischel, 1968; Epstein, 1979, 1983; Fleeson, 2001; Fleeson and Noftle, 2008).

(3) As observed above, EMA provides a means of repeatedly and more directly assessing dispositional expression across an individual's daily situations. It, thus, provides a vehicle for not

FIGURE 2 | Illustrative frequency distributions of momentary, cross-situational dispositional expressions displaying differences in degree of virtue possession. (A) The red curve displays a distribution representing a higher typical degree of virtue expression (higher mean score) than the blue curve. (B) The red curve represents a higher typical degree of (higher mean score), and more consistent (less variable), virtue expression than the blue curve.

only assessing whether an individual has a virtue, but also the degree to which they typically express a virtue and the consistency with which they express a virtue. And since the degree to which an individual typically expresses a virtue and the consistency with which they express the virtue have implications for flourishing, EMA provides a thorough means of examining the relationship between virtues and flourishing.

To illustrate, EMA opens up a way of measuring *the degree to which an individual is typically grateful* and *the consistency with which an individual is grateful*. And both should be expected to promote flourishing given gratitude is a virtue. Thus, by allowing a direct assessment of both, EMA provides a thorough and direct means of examining the relationship between dispositional gratitude and flourishing. Further, since EMA allows the detection of dispositional expression within daily life, it also provides a good means of examining what may mediate relationships between dispositions, like gratitude, and flourishing as well as the *interconnectedness* of virtues (e.g., whether developing dispositional gratitude might directly correlate with developing others).

Before continuing we should mention that there are several limitations associated with EMA. Asking participants to repeatedly respond to prompts and questions over time, and within daily life, places a high demand on participants thereby increasing the likelihood of participant dropout and decreasing response rates (e.g., Shiffman et al., 2008). To compensate,

participants need greater incentive than with traditional surveys. Also, there are issues regarding the invasion of privacy, which must be carefully addressed (Trull, 2015). Further, under certain conditions, EMA has been shown to result in reactivity (cf. Shiffman et al., 2008).

Having examined the advantages and limitations of using EMA—and specifically app-based EMA—to study virtues, we will now examine recent developments in EMI and the possibilities they open up for promoting dispositional development.

EMI, Positive Change, and Smartphone Apps

Clinicians and therapists have often sought ways to improve the impact of therapy between sessions and the efficacy of interventions (Heron and Smyth, 2010). With this aim, over the past several years, researchers have been exploring the use of mobile devices to intervene and interact with clients within the context and moments of their daily life. This form of intervention, called ecological momentary intervention (EMI), has—similar to EMA—been implemented using PDAs, phone calls, text messages and, most recently, smartphone apps. EMI has developed largely as an extension of interventions involving computer- and internet-based cognitive-behavioral therapy (CBT; e.g., Butler et al., 2006; Andersson and Cuijper, 2009; Moore et al., 2011; Spence et al., 2011).

Self-monitoring has long been known to, under certain conditions, raise self-awareness and promote positive behavioral development (e.g., Harris and Lahey, 1982; Korotitsch and Nelson-Gray, 1999; Shapiro and Cole, 1999; Shiffman et al., 2008; Cohen et al., 2013; Maas et al., 2013). Specifically, it has been theorized that being asked questions about one's momentary states, experiences, behaviors and/or thoughts close to the time and context of their occurrence may help one become more mindful of their occurrence thereby providing opportunity for change (cf. Goodwin et al., 2008)[5]. Recent evidence suggests that EMI may be particularly effective for self-monitoring and raising self-awareness (Robinson et al., 2013; Runyan et al., 2013). For instance, in a recent study, our lab used an EMA/I app (iHabit) to ask undergraduate freshman how they were spending their time at various points throughout the day for three separate weeks during a semester (Runyan et al., 2013). Compared to controls, at the end of the study freshman using the app reported wasting nearly twice as much time throughout the semester. Further, amongst those using the app—but not amongst controls—this self-report predicted semester GPA comparable to the best single predictors of first semester GPA (i.e., high-school GPA and

---

[5]It should be noted that it has also been theorized that simply being aware of being observed, or of having one's behavior assessed, effects one's behavior—a phenomenon known as the Hawthorne effect (Solomon, 1949; Sommer, 1968; Parsons, 1974; Wickstrom and Bendix, 2000). However, recently, mixed evidence for the Hawthorne effect has lead to serious questions about the adequacy of this construct. Rather than there being a Hawthorne effect, evidence indicates that multiple conditions lead to various effects when one is aware of being assessed or of having one's behavior assessed (see Gale, 2004; McCambridge et al., 2014). Here, we focus on one such condition: increased self-awareness.

ACT score). The implication is that using the app promoted self-awareness concerning time-management.

In addition to prompting self-monitoring, EMIs have been used to:

- motivate (Rodgers et al., 2005; Franklin et al., 2006);
- encourage engagement in practices or the use of previously learned skills (e.g., DBT coach, Rizvi et al., 2011);
- aid in the development of new skills (Villani et al., 2012; Bless et al., 2014);
- notify (Dulin et al., 2014) or distract (Rodgers et al., 2005) individuals when they are at risk of engaging in addictive behavior; and
- provide individuals with personalized summary data (Hurling et al., 2007; Bless et al., 2014).

Though most EMI studies using smartphone apps are in developmental stages (e.g., Pramana et al., 2014; Wenze et al., 2014), there is some initial evidence from a number of health, clinical, and therapeutic domains that, in certain forms, EMI may promote positive dispositional development (e.g., Heron and Smyth, 2010; Cohn et al., 2011; Donker et al., 2013; see **Table 5**).

For example, in a smoking cessation study, text messages were sent to participants about health practices. Participants were also sent motivational stories or distraction topics (sports, travel, general interest, etc.) during times they were likely to smoke (Rodgers et al., 2005). Nearly 1000 messages were designed for this study and were sent to individual participants based on factors such as smoking history and preferences. The intervention was largely successful. In comparison to controls, twice as many people in the EMI group reported that they had quit smoking after 6 weeks.

In another recent EMI study, participants performed a cognitive task to improve auditory attention on an iPod touch twice a day for 3 weeks (Bless et al., 2014). This experimental group, unlike the control group, showed improved auditory attention and evidence of functional neural plasticity in regions associated with auditory processing (left posterior temporal gyrus) and executive function (right middle frontal gyrus) during an auditory attention task as measured using fMRI (see **Figure 3**).

As we noted earlier, an important part of the psychological study of virtue is the examination of whether stable and robust dispositions can be developed. And though most app-based EMIs are in various exploratory stages, as we will discuss next, they hold promise for incorporating interventions that promote positive dispositional development into the daily activities of a large non-clinical, non-therapeutic population. In initial studies, effects during and immediately following EMI have been documented more than long-term effects (see **Table 5**). However, smartphone app-based systems that couple EMA and EMI provide an unprecedented means of testing whether EMI can help promote dispositional, including virtue, development.

Virtues and App-Based EMI

Little is directly known about the efficacy of EMI approaches to virtue development. In this section we, thus, discuss reasons for thinking such approaches have promise. In particular, we point out how app-based EMI offers a versatile, multifaceted and interactive way of promoting training, mindfulness, self-awareness, motivation and environmental awareness within the *context* of everyday life. We, then, outline parameters that may influence the effectiveness of EMI and an approach to optimizing EMIs for virtue development.

In addition to being practical due to the widespread use of smartphones, app-based EMI may be a particularly efficacious approach to promoting virtue development since it provides a versatile and multifaceted means of interacting with individual's within their everyday context. Developing a disposition, such as a virtue, is a learning process through which a behavior or response becomes a stable and robust habitual response or automatic
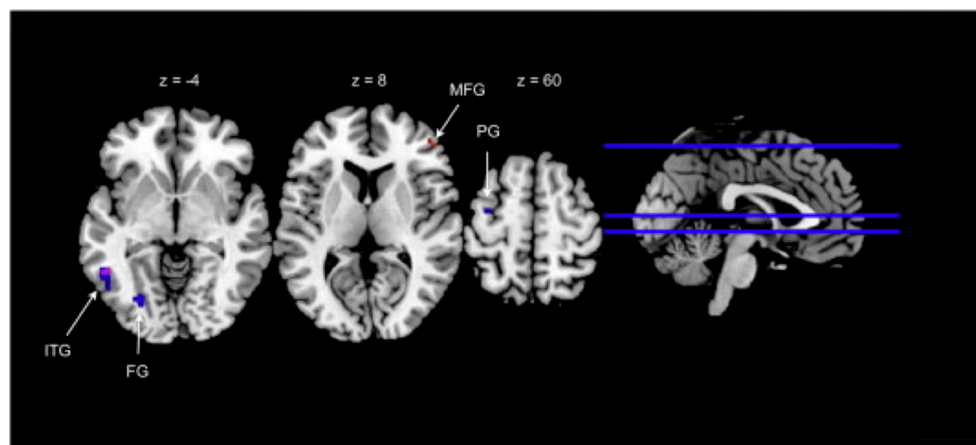
FIGURE 3 | Neural differences between those trained in an auditory attention task and controls as measured by fMRI. Brain regions displaying significant decreases at points during a selective auditory attention task amongst trained participants (taken and adapted from Bless et al., 2014). Z, horizontal plane coordinate; ITG, inferior temporal gyrus; FG, fusiform gyrus; PG, precentral gyrus; MFG, middle frontal gyrus; red, forced-left response conditions; blue, forced-right response conditions; purple, overlap.

TABLE 5 | Peer-reviewed studies reporting effective smartphone/iPod EMIs.

Study	Intervention target	Sensors	Tailoring	App-based	EMI frequency	EMI duration	Immediate Post-intervention effects	Long-term effects
Obermayer et al., 2004	Smoking cessation		×		Variant; At least 2 per day; Combined with website	6 weeks	x[a]	
Rodgers et al., 2005	Smoking cessation		×		5 per day for 6 weeks; 3 per week thereafter	6 months	×	×
Franklin et al., 2006	Type 1 diabetes management		×		1 daily and 1 weekly message	12 months	×	
Kim and Jeong, 2007	Type 2 diabetes management		×		Weekly	6 months	×	
Joo and Kim, 2007	Weight management behavior				Weekly	12 weeks	×	
Hurling et al., 2007	Physical activity	×	×		Tailored; Combined with website and email	4 months	×	
Weitzel et al., 2007	Negative consequences of drinking		×		Daily surveys and messages	2 weeks	×	
Yoon and Kim, 2008	Type 2 diabetes management		×		Weekly	12 months	×	
Brendryen et al., 2008	Smoking cessation				Between 1 and 3 messages every 2 weeks; Combined with website, interview and email	12 months	×	×
Atienza et al., 2008	Dietary intake		×		2 per day	8 weeks	×	
King et al., 2008	Physical activity		×		2 daily assessments; 1 daily & weekly tailored	8 weeks	×	
Patrick et al., 2009	Weight loss		×		Variant; Typically 1–5 per day	4 months	×	
Rizvi et al., 2011	Borderline personality disorder and substance use disorder			×	On average 15 times total	10–14 days	×	
Burns et al., 2011	Major depressive disorder	×		×	Tailored	8 weeks	x[a]	
Pop-Eleches et al., 2011	Antiretroviral therapy adherence				Weekly	48 weeks	×	×
Granholm et al., 2012	Schizophrenia-related medication adherence, socialization and hallucinations				12 per day; Combined with initial training	12 weeks	×	
Villani et al., 2012	Stress			×	8 videos	4 weeks	×	
Kauer et al., 2012	Emotional self-awareness and depressive symptoms			×	Approximately 2 per day (self-initiated)	2–4 weeks		×
Watts et al., 2013	Depression		×		6 sessions; Combined with weekly assignments and email	n/a	×	×
Robinson et al., 2013	Eating behaviors and weight management		×	×	2.7 episodes on average (user-initiated)	27.5 days on average	x[b]	
Carta et al., 2013	Parenting strategies and child engagement				2 per day; Combined with training home visits and weekly phone calls	Not specified	×	×
King et al., 2013	Motivation and physical activity	×	×	×	Live wallpaper and 1 end of day intervention; Combined with user-initiated features, texts and weekly alerts if goals weren't met	8 weeks	×	×

(Continued)

TABLE 5 | Continued

Study	Intervention target	Sensors	Tailoring	App-based	EMI frequency	EMI duration	Immediate Post-intervention effects	Long-term effects
Wayne and Ritvo, 2014	Self-management of type 2 diabetes	×		×	User-initiated; Combined with phone calls and in person meetings	6 months	×	
Enock et al., 2014	Social anxiety		×		3 per day	4 weeks	×	
Bless et al., 2014	Auditory attention		×	×	2 per day	21 days	×	
Ben-Zeev et al., 2014	Schizophrenia			×	5.2 per day on average	1 month	×	
Bond et al., 2014	Sedentary time reduction	×	×	×	Dependent on activity levels; Combined with education session	3 weeks	×	
Dulin et al., 2014	Treatment of alcohol use disorders	×	×	×	Variant; Location-based	6 weeks	×	
Lane et al., 2014	Sleep, physical activity, social interaction	×	×	×	Constantly accessible summary data	19 days	×[c]	
Maddison et al., 2014	Self-efficacy and exercise				Variant; First 12-weeks more intense; Combined with website	2 months	×	
Mhurchu et al., 2014	Weight management behavior	×	×		Ave. 2 motivational texts per day; 2 self-monitoring texts per week; Combined with hard-copy "toolkit" and website	8 weeks; 4 week maintenance		×[d]
Kramer et al., 2014	Depression		×		10 per day for 3 consecutive days per week; Weekly face-to-face feedback sessions	6 weeks	×	×
Cranwell et al., 2014	Self-control				3 per day	4 weeks	×	
Macias et al., 2015	Psychiatric and physical well-being	×		×	1 semiweekly, 1 weekly and 1 end of study assessment; Combined with texts, digital readings and videos	4 weeks	×	×

These studies were located by searching Pubmed.com, and publications from peer-reviewed journals listed in Scholar.google.com, up to November 27th, 2014 (search terms: "ecological momentary intervention" and "smartphone").
[a]*Did not use controls.*
[b]*Half of participants lost weight. However, there was no control group as this was an intervention feasibility study.*
[c]*These results are from a small pilot, feasibility study.*
[d]*A pre/post decrease in weight and body mass index (BMI) was observed but there was no control as this was mainly a feasibility study.*

response (e.g., Wood and Neal, 2007; Gawronski and Cesario, 2013). Learning context is important for this process.

There has been extensive animal research on the importance of learning context for response learning. First, animal studies reveal that learning contexts can function as "occasion-setters" such that, while they do not elicit a learned response themselves, they influence an animal's learned response to another stimulus thereby "setting the occasion" for this response (cf. Schmajuk and Holland, 1998; Bouton, 2010). Second, after learning a new response, animals often revert back to previous responses within contexts that vary from the context in which the new response was learned (cf. Bouton and Bolles, 1979; Peck and Bouton, 1990). Third, animals can be conditioned in one context (context A) and counterconditioned—or conditioned to give an opposing response—in another (context B), and continue to show the initial, conditioned response in context A and the opposing, counterconditioned response in context B (cf. Bouton and Bolles, 1979; Bouton and Peck, 1989; Merchant et al., 2013). Fourth, it has been observed that, under certain training conditions, a learned contextual response can persist even though memory for a conditioned stimulus (CS)/unconditioned stimulus (US) association has been impaired through localized inhibition of neurobiochemical processes crucial for the formation of long-term memory (i.e., memory persisting at least 48 h) for CS/US association (Runyan et al., 2004).

Extending from these animal studies, human studies have revealed that context is similarly important in the learning of habitual responses and automatic responses (e.g., Rydell and Gawronski, 2009; Wood and Neal, 2009; Gawronski and Cesario, 2013). Additionally, there is evidence that contextual cues can influence preparatory neural states associated with a certain response and the likelihood an individual will respond in that way in certain contexts (e.g., Deiber et al., 1996; Thoenissen et al., 2002; Toni et al., 2002; Praamstra et al., 2009; Moisa et al., 2012). In particular, addiction studies have shown that recovering individual's are more likely to relapse within contexts associated with the addictive behavior (e.g., Crombag et al., 2008); and this has been associated with plasticity in specific brain regions, including the lateral hypothalamus (e.g., Marchant et al., 2009, 2014). Further, there is evidence indicating that the tendency to give a habitual response becomes more *stable* and *robust* with the repetition of the response in various contexts (e.g., Bouton, 2000; Neal et al., 2006, 2011). Taken together, these observations provide evidence that, by promoting the development of habitual responses or automatic responses within an individual's daily context, EMIs aimed at virtue development may be particularly effective. Given this, EMA data coupled with GPS data could be used to trigger app-based EMIs in particular spatiotemporal locations to promote the learning of habitual or automatic responses in new, or various, contexts; especially in contexts where an individual finds change difficult (e.g., Watkins et al., 2014).

One way that EMI might be effective in promoting virtue development is by prompting individuals to engage in practices, or in training, aimed at developing a particular virtue (cf. Magidson et al., 2014). Working memory training has been shown to improve cognitive abilities and to result in neural

plasticity (e.g., Klingberg, 2010). There is some suggestion that practicing gratitude increases dispositional gratitude (Emmons and McCullough, 2003; Seligman et al., 2005). Additionally, self-regulation exercises have been shown to improve self-regulation (Baumeister et al., 2006; Cranwell et al., 2014). Further, in a recent study, participants were given empathy and compassion training (Klimecki et al., 2014; cf. Klimecki et al., 2012). After empathy training, participants experienced increased negative affect associated with increased activity in the anterior insula and the anterior midcingulate cortex (two regions previously associated with empathy for pain) in response to watching videos depicting human suffering. After compassion training, these same individuals also experienced increased positive affect associated with increased activity in the ventral striatum, pregenual anterior cingulate cortex and medial orbitofrontal cortex (see **Figure 4**). The same effects were not found in controls who underwent memory training. The implication is that training can result in increased empathy and compassion as measured by affective responses and associated brain activity states.

App-based EMI may be a particularly effective way of administering dispositional training of the kind mentioned above since smartphone app technology can support multifaceted, interactive and progressive training within various contexts throughout an individual's daily routine. For instance, training might involve prompting individuals to engage in app-delivered activities or exercises, interact with app-based games, or with videos or pictures, throughout the day. And training might get more demanding over time.

Another way EMI might promote virtue development is by specifically targeting mindfulness. According to the predominating definition in psychology, mindfulness is purposeful, non-evaluative awareness of one's present experiences and mental states moment-to-moment (Kabat-Zinn, 2003; but see Brown and Ryan, 2003; Jankowski and Holas, 2014). Not only is mindfulness thought by some to be a virtue—and there is evidence that dispositional mindfulness does promote mental health and resilience (e.g., Brown et al., 2007)—there is some suggestion that mindfulness may help promote the development of other virtues and, thus, provide a case where virtues are interconnected. It is theorized that mindfulness promotes positive change by promoting awareness of one's immediate experiences and states from a somewhat detached state, which may, in turn, promote self-awareness and self-regulation (Shapiro et al., 2006; Jankowski and Holas, 2014).

EMA observations support the theory that mindfulness promotes self-regulation (e.g., Brown and Ryan, 2003). Additionally, mindfulness training (involving attentional fixation on and nonjudgmental awareness of moment-to-moment experiences) can improve working memory and attention (Tang et al., 2007; Lutz et al., 2008; Jha et al., 2010; MacLean et al., 2010). These improvements have been associated with increased activity in the left dorsolateral prefrontal and dorsal anterior cingulate cortex (Allen et al., 2012). Mindfulness has also been observed to reduce interference from emotionally salient distractors (Ortner et al., 2007),
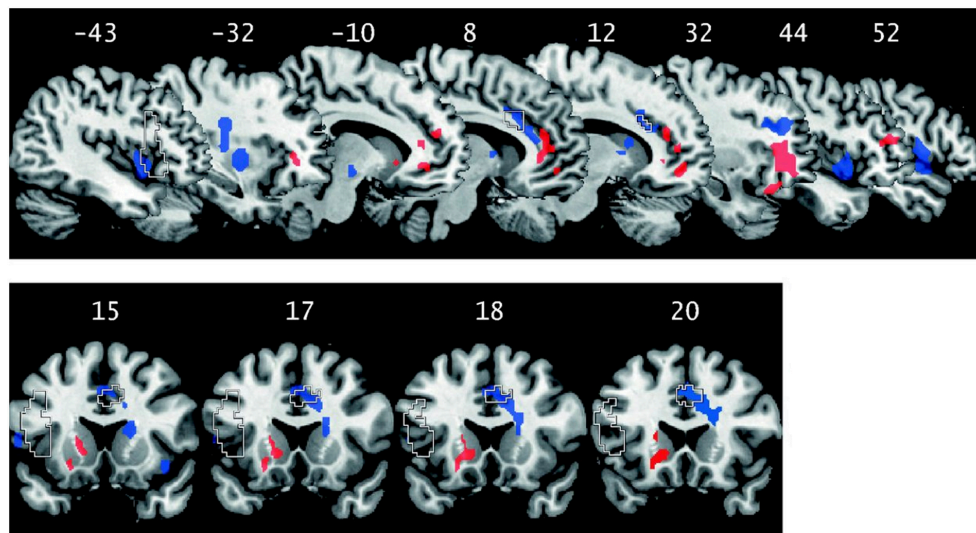
FIGURE 4 | Functional neural changes corresponding with empathy training (red) and compassion training (blue) in comparison to memory control as measured by fMRI (adapted from Klimecki et al., 2014). Dashed lines represent regions associated with empathy for pain in a recent meta-analysis by Lamm et al. (2011). Numbers on the top row are the sagittal plane coordinates. Numbers on the bottom row are the coronal plane coordinates.

which has been associated with increased activity in the medial prefrontal cortex and right anterior insula (Allen et al., 2012).

App-based EMI approaches to virtue development might prompt and remind individuals to engage in dynamic and progressive mindfulness exercises within the context of daily life. There are a number of mindfulness apps currently available (Plaza et al., 2013). However, at present there has been little work on the effectiveness of app-based EMIs aimed at promoting mindfulness (but see Chittaro and Vianello, 2014). Further, as we will see toward the end of this section, there are a number of parameters that are likely to influence whether EMIs are efficacious.

In addition to prompting various practices, trainings or exercises within everyday contexts, effective app-based EMIs for virtue development might be designed by incorporating effective components of computer- and internet-based therapeutic interventions (cf. Kaltenthaler et al., 2006; Barak et al., 2008; Tillfors et al., 2008; Andersson, 2009; Bergström et al., 2010; Newman et al., 2011; Andersson et al., 2013; Musiat and Tarrier, 2014). Web-based cognitive-behavioral therapy and guided self-help interventions have been observed to be efficacious in treating eating disorders (Hötzel et al., 2013), insomnia (Holmqvist et al., 2014), suicidal ideation, depressive symptoms, anxiety (van Straten et al., 2008; March et al., 2009; Cuijpers et al., 2013), posttraumatic stress disorder (Knaevelsrud and Maercker, 2007) and physical inactivity (van Stralen et al., 2009a,b, 2010, 2011). In multiple cases, these interventions have been observed to facilitate long-term change (e.g., Knaevelsrud and Maercker, 2007; Litz et al., 2007; March et al., 2009; Ruwaard et al., 2010; van Stralen et al., 2011; Cuijpers et al., 2013; Lappalainen et al., 2014).

In these computer- and internet-based approaches:

(1) self-monitoring and self-awareness (e.g., Litz et al., 2007; Morris et al., 2010),
(2) self-efficacy and motivation (Litz et al., 2007; Turner et al., 2007; Carlbring and Smit, 2008; Warmerdam et al., 2010; Hötzel et al., 2013), and
(3) environmental awareness (van Stralen et al., 2009a)

are among the strongest mediators of enduring change. And, as we will discuss in turn below, app-based EMI might be used to effectively promote each.

(1) As mentioned in the previous section, research from our own lab shows that using an EMA/I app to ask individuals questions about their responses at random times throughout their daily activities can raise self-awareness (Runyan et al., 2013). Additionally, app-based EMIs might allow individuals to monitor summary data generated from their responses. Further, by incorporating sensors, app-based EMI provides a means of increasing self-awareness by increasing a person's ability to self-monitor. App-based EMIs that interface with mobile EEG (cf. Curran and Stokes, 2003), physical activity sensors and/or other physiological activity sensors (e.g., sensors for muscle tension, temperature, galvanic skin response, blood pressure or heart-rate; see Sutarto et al., 2010; Schwerdtfeger and Scheel, 2012; Bossman et al., 2013; von Haaren et al., 2013; Demarble et al., 2014; Dunton et al., 2014) could provide biofeedback allowing an individual to self-monitor to an extent otherwise impossible within close spatial and temporal proximity to a focal event or state (cf. Keedwell and Linden, 2013; Linden, 2014; Schoenberg and David, 2014). Feedback concerning preparatory neural activities, muscle tension, blood pressure and/or heart rate (e.g., Koval et al., 2013) within certain contexts might enable an individual to monitor

patterns in their behaviors, thought life, states and/or experiences of which they would otherwise be unaware.

(2) There is indication that self-monitoring promotes self-regulation and positive behavioral development when individuals have the opportunity and motivation to change (cf. Bandura, 1991; Korotitsch and Nelson-Gray, 1999; Shiffman et al., 2008; Quinn et al., 2010). Therefore, app-based EMIs might effectively promote positive change not only by providing a means for self-monitoring but by also increasing motivation through:

- repeatedly asking individuals directed questions;
- prompting goal-setting;
- pushing motivational statements, videos or pictures and promoting positive self-talk to increase outcome expectancy (i.e., belief that one's efforts will lead to the achievement of a desirable end) and self-efficacy (i.e., belief in one's ability to achieve one's goal; Bandura, 1986, 1997);
- reminding individuals of their own intentions and values; and/or
- intermittently giving individuals progress and accomplishment reports (which has been shown to increase motivation and promote positive behavioral development; e.g., Korotitsch and Nelson-Gray, 1999; Kramer et al., 2014; Mhurchu et al., 2014).

Also, a social component to app-based EMI, where individuals can interact with others who are using the same EMI in order to develop the same virtue, might also increase motivation as well as self-efficacy (cf. Obermayer et al., 2004; Przeworski and Newman, 2004; Hurling et al., 2007; Brendryen et al., 2008; Cafazzo et al., 2012; King et al., 2013).

(3) With regard to increasing environmental awareness, app-based EMIs that incorporate GPS (e.g., Yüce et al., 2012; Hollett and Leander, 2013; Huang and Luo, 2014; Watkins et al., 2014), and utilize smartphone microphones and cameras, might be designed to notify individuals about aspects of their environment. In this way, app-based EMIs might be designed to raise awareness of contextual/situational triggers for responses that the individual desires to change as well as awareness of opportunities to respond in ways that are expressions of a virtue. App-based EMIs might also raise awareness of environmental resources and/or social support that may help an individual in their effort to change (e.g., van Stralen et al., 2009a; Mhurchu et al., 2014).

Whether EMI can effectively promote virtue development—and the optimal conditions for this development—remains to be directly and systematically tested. In particular, it remains to be seen whether EMIs, including app-based EMIs, can promote long-term dispositional development that persists following the termination of the intervention. Relevant parameters of EMI in need of testing are:

(i)   duration of intervention;
(ii) frequency of interactions;
(iii) duration of each interaction;
(iv) whether interactions occur at varied or fixed times;
(v)   whether, and the degree to which, the intervention is intermittent or continuous over a period of time;

(vi)   degree of variation or repetition between interactions;
(vii) how broadly or narrowly the intervention is focused (e.g., number of responses targeted, types of interactions provided);
(viii) whether the intervention is individualized/tailored;
(ix)   whether the intervention progresses over time (e.g., involves scaffolding);
(x)   whether the intervention involves social interaction; and
(xi)   the degree to which the intervention is engaging (e.g., interesting, meaningful and/or fun).

Given EMI can promote virtue development, optimal conditions are likely to depend somewhat on the virtue. Nevertheless, there are also likely to be some common optimal conditions. We hypothesize that these optimal conditions will, to a certain degree, track those for instrumental learning whereby intentionally modified responses in early stages of learning are cued by stimuli as expressions of a habitual response or automatic response in later stages (e.g., Dickinson et al., 1995; Schachtman and Reilly, 2011). In this case, EMIs that maintain the highest levels of engagement, motivation and awareness of opportunities to respond or otherwise behave in ways that are expressions of a virtue, over the longest period of time, are likely to be the most effective (e.g., Rescorla and Solomon, 1967; Sutherland and Mackintosh, 1971; Rescorla and Wagner, 1972; Bandura, 1977). This is likely to be accomplished by EMIs that:

1. focus on one virtue at a time (cf. Kamin, 1968; Rescorla and Wagner, 1972);
2. progress and involve scaffolding (cf. Salomon, 1993; Pea, 2004);
3. are tailored to the individual and incorporate the individual's judgments, values and aims (cf. Levey and Martin, 1975; Shanks and Dickinson, 1991; Kreuter et al., 2000a,b; Strecher et al., 2005);
4. involve social interaction amongst individuals with similar aims (cf. Bandura, 1977, 1986, 1997); and
5. interact with individuals at varied, rather than fixed, intervals and provide positive reinforcement for virtue development (e.g., make virtue development, and interaction with the intervention, varied, interesting, meaningful and/or fun).

Scheduling intermittent EMAs as an individual uses an EMI aimed at promoting virtue development provides a way of assessing and optimizing the efficacy of EMI in real-time (cf. Voogt et al., 2013). As we have already discussed, EMA can be used to assess virtue expression, and/or detect correlates of this expression. And a smartphone app can be used to administer both EMA and EMI.

Conclusions

There seems to be conclusive reasons for thinking that some people possess virtues understood as relatively stable and robust psychological dispositions that contribute to a deeply fulfilling, well-lived life of growth (i.e., a flourishing life)—and situational studies have not presented reasons for thinking otherwise.

Further, since knowledge of what contributes to flourishing is worth seeking, virtues are worth studying. As a result of the widespread use of smartphones and advancements in smartphone technology, app-based EMA provides a new means for examining the stability, robustness and interconnectedness of virtues, and the physiological (including neurophysiological) correlates of having and/or expressing virtues. In short, app-based EMA provides a wholistic approach to examining the extent to which virtues are possessed as well as the personal, physiological and biochemical characteristics of individuals who possess and express virtues, which promises to provide new insight. Additionally, app-based EMI provides a new and powerful vehicle for promoting the development of virtues where virtues should be expressed—within the context of our daily lives.

Acknowledgments

References

aan het Rot, M., Moskowitz, D. S., and Young, S. N. (2015). Impulsive behavior in interpersonal encounters: associations with quarrelsomeness and agreeableness. *Br. J. Psychol.* 106, 151–161. doi: 10.1111/bjop.12070

Ackerman, J. M., Nocera, C. C., and Bargh, J. A. (2010). Incidental haptic sensations influence social judgments and decisions. *Science* 328, 1712–1715. doi: 10.1126/science.1189993

Adams, P., Rabbi, M., Rahman, T., Matthews, M., Voida, A., Gay, G., et al. (2014). "Towards personal stress informatics: comparing minimally invasive techniques for measuring daily stress in the wild," in *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, 72–79.

Allen, M., Dietz, M., Blair, K. S., van Beek, M., Rees, G., Vestergaard-Poulsen, P., et al. (2012). Cognitive-affective neural plasticity following active-controlled mindfulness intervention. *J. Neurosci.* 32, 15601–15610. doi: 10.1523/JNEUROSCI.2957-12.2012

Allport, A. (1960). *Becoming: Considerations for a Psychology of Personality*. New Haven, CT: Yale University Press.

Alzola, M. (2008). Character and environment: the status of virtues in organizations. *J. Bus. Ethics* 78, 343–357. doi: 10.1007/s10551-006-9335-7

Ambrose. (377 AD/1961). *On Paradise*. Transl. J. Savage. Washington, DC: The Catholic University of America Press.

Andersson, G. (2009). Using the internet to provide cognitive behaviour therapy. *Behav. Res. Ther.* 47, 175–180. doi: 10.1016/j.brat.2009.01.010

Andersson, G., and Cuijper, P. (2009). Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cogn. Behav. Ther.* 38, 196–205. doi: 10.1080/16506070903318960

Andersson, G., Hesser, H., Hummerdal, D., Bergman-Nordgren, L., and Calrbring, P. (2013). A 3.5-year follow-up of internet-delivered cognitive behavior therapy for major depression. *J. Mental Health* 22, 155–164. doi: 10.3109/09638237.2011.608747

Anjum, R. L., Lie, S. A. N. and, Mumford, S. (2013). "Dispositions and ethics," in *Powers and Capacities in Philosophy: The New Aristotelianism*, eds R. Groff and J. Greco (New York, NY: Routledge), 231–247.

Annas, J. (2011). *Intelligent Virtue*. New York, NY: Oxford University Press.

Anscombe, G. E. M. (1958). Modern moral philosophy. *Philosophy* 33, 1–19. doi: 10.1017/S0031819100037943

Aquinas, T. (1274/1948). *Summa Theologica*. Transl. Fathers of the Dominican Province. Notre Dame, IN: Christian Classics.

Aristotle. (2000). *The Nicomachean Ethics*. Transl. R. Crisp. Cambridge: Cambridge University Press.

Asch, S. E. (1951). "Effects of group pressure upon the modification and distortion of judgment," in *Groups, Leadership and Men*, ed H. Guetzkow (Pittsburgh, PA: Carnegie Press), 177–190.

Atienza, A. A., King, A. C., Oliveira, B. M., Ahn, D. K., and Gardner, C. D. (2008). Using hand-held computer technologies to improve dietary intake. *Am. J. Prevent. Med.* 34, 514–518. doi: 10.1016/j.amepre.2008.01.034

Baer, R. A. (2003). Mindfulness training as a clinical intervention: a conceptual and empirical review. *Clin. Psychol. Sci. Pract.* 10, 125–143. doi: 10.1093/clipsy.bpg015

Baer, R. A., Lykins, E. L. B., and Peters, J. R. (2012). Mindfulness and self-compassion as predictors of psychological wellbeing in long-term meditators and matched nonmeditators. *J. Posit. Psychol.* 7, 230–238. doi: 10.1080/17439760.2012.674548

Baer, R. A., Smith, G. T., and Allen, K. B. (2004). Assessment of mindfulness by self-report the Kentucky inventory of mindfulness skills. *Assessment* 11, 191–206. doi: 10.1177/1073191104268029

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037/0033-295X.84.2.191

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Org. Behav. Hum. Decis. Process.* 50, 248–287. doi: 10.1016/0749-5978(91)90022-L

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY: Freeman.

Bandura, A. (2004). Health promotion by social cognitive means. *Health Educ. Behav.* 31, 143–164. doi: 10.1177/1090198104263660

Barak, A., Hen, L., Boniel-Nissim, M., and Shapira, N. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *J. Technol. Hum. Serv.* 26, 109–160. doi: 10.1080/15228830802094429

Bargh, J. A., and Shalev, I. (2012). The substitutability of physical and social warmth in daily life. *Emotion* 12, 154–162. doi: 10.1037/a0023527

Bartlett, M. Y., and DeSteno, D. (2006). Gratitude and prosocial behavior helping when it costs you. *Psychol. Sci.* 17, 319–325. doi: 10.1111/j.1467-9280.2006.01705.x

Baumeister, R. F., Gailliot, M., DeWall, C. N., and Oaten, M. (2006). Self−regulation and personality: how interventions increase regulatory success, and how depletion moderates the effects of traits on behavior. *J. Personal.* 74, 1773–1802. doi: 10.1111/j.1467-6494.2006.00428.x

Ben-Zeev, D., Brenner, C. J., Begale, M., Duffecy, J., Mohr, D. C., Mueser, K. T. (2014). Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophr. Bull.* 40, 1244–1253. doi: 10.1093/schbul/sbu033

Bergström, J., Andersson, G., Ljótsson, B., Rück, C., Andréewitch, S., Karlsoon, A., et al. (2010). Internet-versus group-administered cognitive behaviour therapy for panic disorder in a psychiatric setting: a randomised trial. *BMC Psychiatry* 10:54. doi: 10.1186/1471-244X-10-54

Bernstein, A., Tanay, G., and Vujanovic, A. A. (2011). Concurrent relations between mindful attention and awareness and psychopathology among trauma-exposed adults: preliminary evidence of transdiagnostic resilience. *J. Cogn. Psychother.* 25, 99–113. doi: 10.1891/0889-8391.25.2.99

Berry, J. W., and Worthington, E. L. Jr. (2001). Forgivingness, relationship quality, stress while imagining relationship events, and physical and mental health. *J. Couns. Psychol.* 48, 447–455. doi: 10.1037/0022-0167.48.4.447

Bleidorn, W. (2009). Linking personality states, current social roles and major life goals. *Eur. J. Pers.* 23, 509–530. doi: 10.1002/per.731

Bleidorn, W., and Denissen, J. J. A. (2015). Virtues in action—the new look of character traits. *Br. J. Psychol.* doi: 10.1111/bjop.12117. [Epub ahead of print].

Bless, J. J., Westerhausen, R., Kompus, K., Gudmundsen, M., and Hugdahl, K. (2014). Self-supervised, mobile-application based cognitive training of auditory attention: a behavioral and fMRI evaluation. *Inter. Intervent.* 1, 102–110. doi: 10.1016/j.invent.2014.06.001

Bond, D. S., Thomas, J. G., Raynor, H. A., Moon, J., Sieling, J., Trautvetter, J., et al. (2014). B-MOBILE-A smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. *PLoS ONE* 9:e100821. doi: 10.1371/journal.pone.0100821

Bossman, T., Kanning, M., Koudela-Hamila, S., Hey, S., and Ebner-Priemer, U. (2013). The association between short periods of everyday life activities and affective states: a replication study using ambulatory assessment. *Front. Psychol.* 4:102. doi: 10.3389/fpsyg.2013.00102

Bouton, M. E. (2000). A learning theory perspective on lapse, relapse, and the maintenance of behavioral change. *Health Psychol.* 19, 57–63. doi: 10.1037/0278-6133.19.Suppl1.57

Bouton, M. E. (2010). "The multiple forms of 'context' in associative learning theory," in *The Mind in Context*, eds B. Mesquita, L. Feldman, and E. R. Smith (New York, NY: Guilford Press), 233–258.

Bouton, M. E., and Bolles, R. C. (1979). Role of conditioned contextual stimuli in reinstatement of extinguished fear. *J. Exp. Psychol. Anim. Behav. Process.* 5, 368–378. doi: 10.1037/0097-7403.5.4.368

Bouton, M. E., and Peck, C. A. (1989). Context effects on conditioning, extinction, and reinstatement in an appetitive conditioning preparation. *Anim. Learn. Behav.* 17, 188–198. doi: 10.3758/BF03207634

Bowlin, S. L., and Baer, R. A. (2012). Relationships between mindfulness, self-control, and psychological functioning. *Pers. Indiv. Diff.* 52, 411–415. doi: 10.1016/j.paid.2011.10.050

Brendryen, H., Drozd, F., and Kraft, P. (2008). A digital smoking cessation program delivered through internet and cell phone without nicotine replacement (happy ending): randomized controlled trial. *J. Med. Inter. Res.* 10, e51. doi: 10.2196/jmir.1005

Brown, G. L., McGarvey, E. L., Shirtcliff, E. A., Keller, A., Granger, D. A., and Flavin, K. (2008). Salivary cortisol, dehydroepiandrosterone, and testosterone interrelationships in healthy young males: a pilot study with implications for studies of aggressive behavior. *Psychiatry Res.* 159, 67–76. doi: 10.1016/j.psychres.2007.06.012

Brown, K. W., and Ryan, R. M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *J. Pers. Soc. Psychol.* 84, 822–848. doi: 10.1037/0022-3514.84.4.822

Brown, K. W., Ryan, R. M., and Creswell, J. D. (2007). Mindfulness: theoretical foundations and evidence for its salutary effects. *Psychol. Inquiry* 18, 211–237. doi: 10.1080/10478400701598298

Brown, S. L., Smith, D. M., Schulz, R., Kabeto, M. U., Ubel, P. A., Poulin, M., et al. (2009). Caregiving behavior is associated with decreased mortality risk. *Psychol. Sci.* 20, 488–494. doi: 10.1111/j.1467-9280.2009.02323.x

Bruehl, S., Liu, X., Burns, J. W., Chont, M., and Jamison, R. N. (2012). Associations between daily chronic pain intensity, daily anger expression, and trait anger expressiveness: an ecological momentary assessment study. *Pain* 153, 2352–2358. doi: 10.1016/j.pain.2012.08.001

Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., et al. (2011). Harnessing context sensing to develop a mobile intervention for depression. *J. Med. Inter. Res.* 13, e55. doi: 10.2196/jmir.1838

Burt, S. A., and Donnellan, M. B. (2010). Evidence that the subtypes of antisocial behavior questionnaire (STAB) predicts momentary reports of acting-out behavior. *Pers. Indiv. Diff.* 48, 917–920. doi: 10.1016/j.paid.2010.02.021

Butler, A. C., Chapman, J. E., Forman, E. M., and Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin. Psychol. Rev.* 26, 17–31. doi: 10.1016/j.cpr.2005.07.003

Cafazzo, J. A., Casselman, M., Hamming, N., Katzman, D. K., and Palmert, M. R. (2012). Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. *J. Med. Inter. Res.* 14, e70. doi: 10.2196/jmir.2058

Carlbring, P., and Smit, F. (2008). Randomized trial of internet-delivered self-help with telephone support for pathological gamblers. *J. Consult. Clin. Psychol.* 76, 1090–1094. doi: 10.1037/a0013603

Carta, J. J., Lefever, J. B., Bigelow, K., Borkowski, J., and Warren, S. F. (2013). Randomized trial of a cellular phone-enhanced home visitation parenting intervention. *Pediatrics* 132, S167–S173. doi: 10.1542/peds.2013-1021Q

Carver, C. S., and Scheier, M. F. (2014). Dispositional optimism. *Trends Cogn. Sci.* 18, 293–299. doi: 10.1016/j.tics.2014.02.003

Chittaro, L., and Vianello, A. (2014). Computer-supported mindfulness: evaluation of a mobile thought distancing application on naive meditators. *Int. J. Hum. Comput. Stud.* 72, 337–348. doi: 10.1016/j.ijhcs.2013.11.001

Cohen, J. S., Edmunds, J. M., Brodman, D. M., Benjamin, C. L., and Kendall, P. C. (2013). Using self-monitoring: implementation of collaborative empiricism

in cognitive-behavioral therapy. *Cogn. Behav. Pract.* 20, 419–428. doi: 10.1016/j.cbpra.2012.06.002

Cohn, A. M., Hunter–Reel, D., Hagman, B. T., and Mitchell, J. (2011). Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. *Alcohol. Clin. Exp. Res.* 35, 2209–2215. doi: 10.1111/j.1530-0277.2011.01571.x

Colby, A., and Damon, W. (1992). *Some Do Care: Contemporary Lives of Moral Commitment.* New York, NY: The Free Press.

Colby, A., and Damon, W. (1999). "The development of extraordinary moral commitment," in *Morality in Everyday Life: Developmental Perspectives*, eds M. Killen and D. Hart (Cambridge: Cambridge University Press), 342–370.

Coles, M. G. H. (1989). Modern mind–brain reading: psychophysiology, physiology, and cognition. *Psychophysiology* 26, 251–269. doi: 10.1111/j.1469-8986.1989.tb01916.x

Coles, M. G. H., Gratton, G., and Donchin, E. (1988). Detecting early communication: using measures of movement-related potentials to illuminate human information processing. *Biol. Psychol.* 26, 69–89. doi: 10.1016/0301-0511(88)90014-2

Conner, T., and Barrett, L. S. (2005). Implicit self-attitudes predict spontaneous affect in daily life. *Emotion* 5, 476–488. doi: 10.1037/1528-3542.5.4.476

Conner, T. S. (2014). *Experience Sampling and Ecological Momentary Assessment with Mobile Phones.* Available online at: http://www.otago.ac.nz/psychology/otago047475.pdf

Conway, J. M., Rogelberg, S. G., and Pitts, V. E. (2009). Workplace helping: interactive effects of personality and momentary positive affect. *Hum. Perform.* 22, 321–339. doi: 10.1080/08959280903120279

Cranwell, J., Benford, S., Houghton, R. J., Golembewski, M., Fischer, J. E., and Hagger, M. S. (2014). Increasing self-regulatory energy using an internet-based training application delivered by smartphone technology. *Cyberpsychol. Behav. Soc. Netw.* 17, 181–186. doi: 10.1089/cyber.2013.0105

Crombag, H. S., Bossert, J. M., Koya, E., and Shaham, Y. (2008). Context-induced relapse to drug seeking: a review. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 3233–3243. doi: 10.1098/rstb.2008.0090

Croom, A. M. (2014). Vindicating virtue: a critical analysis of the situationist challenge against Aristotelian moral psychology. *Integr. Psychol. Behav. Sci.* 48, 18–47. doi: 10.1007/s12124-013-9249-8

Csikszentmihalyi, M., and Larson, R. (1987). Validity and reliability of the experience-sampling method. *J. Nervous Mental Dis.* 175, 526–536. doi: 10.1097/00005053-198709000-00004

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., and Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Can. J. Psychiatry* 58, 386–392.

Curran, E. A., and Stokes, M. A. (2003). Learning to control brain activity: a review of the production and control of EEG components for driving brain–computer interface (BCI) systems. *Brain Cogn.* 51, 326–336. doi: 10.1016/S0278-2626(03)00036-8

D'Antono, B., Ditto, B., Moskowitz, D. S., and Rios, N. (2001). Interpersonal behavior and resting blood pressure in college women: a daily monitoring study. *J. Psychosom. Res.* 50, 309–318. doi: 10.1016/S0022-3999(01)00199-4

Dahlsgaard, K., Peterson, C., and Seligman, M. E. P. (2005). Shared virtue: the convergence of valued human strengths across culture and history. *Rev. Gen. Psychol.* 9, 203–213. doi: 10.1037/1089-2680.9.3.203

Darley, J. M., and Batson, C. D. (1973). "From Jerusalem to Jericho": a study of situational and dispositional variables in helping behavior. *J. Pers. Soc. Psychol.* 27, 100–108. doi: 10.1037/h0034449

Deci, E. L., and Ryan, R. M. (2008a). Hedonia, eudaimonia, and well-being: an introduction. *J. Happ. Stud.* 9, 1–11. doi: 10.1007/s10902-006-9018-1

Deci, E. L., and Ryan, R. M. (2008b). Facilitating optimal motivation and psychological well-being across life's domains. *Can. Psychol.* 49, 14–23. doi: 10.1037/0708-5591.49.1.14

Deecke, L., Lang, W., Uhl, F., and Podreka, I. (1990). "Looking where the action is: negative DC shifts as indicators of cortical activity," in *From Neuron to Action: An Appraisal of Fundamental and Clinical Research*, eds L. Deecke, J. C. Eccles, and V. Mountcastle (Berlin; Heidelberg: Springer), 25–41.

Deiber, M. P., Ibañez, V., Sadato, N., and Hallett, M. (1996). Cerebral structures participating in motor preparation in humans: a positron emission tomography study. *J. Neurophysiol.* 75, 233–247.

Demarble, J. B., Moskowitz, D. S., Tardif, J.-C., and D'Antono, B. (2014). The relation between hostility and concurrent levels of inflammation is sex, age, and measure dependent. *J. Psychosom. Res.* 76, 384–393. doi: 10.1016/j.jpsychores.2014.02.010

DeSteno, D., Bartlett, M. Y., Baumann, J., Williams, L. A., and Dickens, L. (2010). Gratitude as moral sentiment: emotion-guided cooperation in economic exchange. *Emotion* 10, 289–293. doi: 10.1037/a0017883

Dickinson, A., Balleine, B., Watt, A., Ganzalez, F., and Boakes, R. A. (1995). Motivational control after extended instrumental training. *Anim. Learn. Behav.* 23, 197–206. doi: 10.3758/BF03199935

Diener, E., and Lucas R. E. (1999). "Personality and subjective well-being," in *Well-Being: The Foundations of Hedonic Psychology*, eds D. Kahneman, E. Diener, and N. Schwartz (New York, NY: Sage), 213–229.

Diener, E., Lucas, R. E., and Scollon, C. N. (2006). Beyond the hedonic treadmill: revising the adaptation theory of well-being. *Am. Psychol.* 61, 305–314. doi: 10.1037/0003-066X.61.4.305

Dlugokinski, E., and Firestone, I. J. (1973). Congruence among four methods of measuring other-centeredness. *Child Dev.* 44, 304–308. doi: 10.2307/1128051

Dlugokinski, E., and Firestone, I. J. (1974). Other centeredness and susceptibility to charitable appeals: effects of perceived discipline. *Dev. Psychol.* 10, 21–28. doi: 10.1037/h0035560

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., and Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *J. Med. Inter. Res.* 15, e247. doi: 10.2196/jmir.2791

Donnellan, M. B., and Lucas, R. E. (2009). Introduction to personality and assessment at age 40: reflections on the legacy of the person-situation debate and the future of person-situation integration. *J. Res. Pers.* 43, 117–119. doi: 10.1016/j.jrp.2009.02.010

Doris, J. M. (1998). Persons, situations, and virtue ethics. *Nous* 32, 504–530. doi: 10.1111/0029-4624.00136

Doris, J. M. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Dulin, P. L., Gonzalez, V. M., and Campbell, K. (2014). Results of a pilot test of a self-administered smartphone-based treatment system for alcohol use disorders: usability and early outcomes. *Subst. Abus.* 35, 168–175. doi: 10.1080/08897077.2013.821437

Dunlop, W. L., and Walker, L. J. (2013). The personality profile of brave exemplars: a person-centered analysis. *J. Res. Pers.* 47, 380–384. doi: 10.1016/j.jrp.2013.03.004

Dunton, G. F., Dzubur, E., Kawabata, K., Yanez, B., Bo, B., and Intille, S. (2014). Development of a smartphone application to measure physical activity using sensor-assisted self-report. *Front. Public Health* 2:12. doi: 10.3389/fpubh.2014.00012

Dunton, G. F., Huh, J., Leventhal, A. M., Riggs, N., Hedeker, D., Spruijt-Metz, D., et al. (2013). Momentary assessment of affect, physical feeling states, and physical activity in children. *Health Psychol.* 33, 255–263. doi: 10.1037/a0032640

Edmondson, D., Shaffer, J. A., Chaplin, W. F., Burg, M. M., Stone, A. A., and Schwartz, J. E. (2013). Trait anxiety and trait anger measured by ecological momentary assessment and their correspondence with traditional trait questionnaires. *J. Res. Pers.* 47, 843–852. doi: 10.1016/j.jrp.2013.08.005

eMarketer. (2015). *2 Billion Consumers Worldwide to Get Smart(Phones) by 2016.* Available online at: http://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694

Emmons, R. A., and McCullough, M. E. (2003). Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. *J. Pers. Soc. Psychol.* 84, 377–389. doi: 10.1037/0022-3514.84.2.377

Emmons, R. A., and Mishra, A. (2011). "Why gratitude enhances well-being: what we know, what we need to know," in *Designing Positive Psychology: Taking Stock and Moving Forward*, eds K. M. Sheldon, T. B. Kashdan, and M. F. Steger (New York: Oxford University Press), 248–262.

Engberg, H., Jeune, B., Andersen-Ranberg, K., Martinussen, R., Vaupel, J. W., and Christensen, K. (2013). Optimism and survival: does an optimistic outlook predict better survival at advanced ages? A twelve-year follow-up of Danish

nonagenarians. *Aging Clin. Exp. Res.* 25, 517–525. doi: 10.1007/s40520-013-0122-x

Enock, P. M., Hofmann, S. G., and McNally, R. J. (2014). Attention bias modification training via smartphone to reduce social anxiety: a randomized, controlled multi-session experiment. *Cogn. Ther. Res.* 38, 200–216. doi: 10.1007/s10608-014-9606-z

Entringer, S., Buss, C., Andersen, J., Chicz-DeMet, A., and Wadhwa, P. D. (2011). Ecological momentary assessment of maternal cortisol profiles over a multiple-day period predicts the length of human gestation. *Psychosom. Med.* 73, 469–474. doi: 10.1097/PSY.0b013e31821fbf9a

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *J. Pers. Soc. Psychol.* 37, 1097–1126. doi: 10.1037/0022-3514.37.7.1097

Epstein, S. (1983). Aggregation and beyond: some basic issues on the prediction of behavior. *J. Pers.* 51, 360–392. doi: 10.1111/j.1467-6494.1983.tb00338.x

Eysenck, H. J., and Eysenck, S. B. G. (1969). *Personality Structure and Measurement.* New York, NY: Routledge.

Farrow, T. F. D., Zheng, Y., Wilkinson, I. D., Spence, S. A., Deakin, J. F. W., Tarrier, N., et al. (2001). Investigating the functional anatomy of empathy and forgiveness. *Neuroreport* 12, 2433–2438. doi: 10.1097/00001756-200108080-00029

Fay, D., and Sonnentag, S. (2012). Within-person fluctuations of proactive behavior: how affect and experienced competence regulate work behavior. *Hum. Perform.* 25, 72–93. doi: 10.1080/08959285.2011.631647

Fleeson, W. (2001). Towards a structure- and process-integrated view of personality: traits as density distributions of states. *J. Pers. Soc. Psychol.* 80, 1011–1027. doi: 10.1037/0022-3514.80.6.1011

Fleeson, W. (2004). Moving personality beyond the person-situation debate: the challenge and the opportunity of within-person variability. *Curr. Dir. Psychol. Sci.* 13, 83–87. doi: 10.1111/j.0963-7214.2004.00280.x

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *J. Pers.* 75, 825–862. doi: 10.1111/j.1467-6494.2007.00458.x

Fleeson, W., and Noftle, E. (2008). The end of the person–situation debate: an emerging synthesis in the answer to the consistency question. *Soc. Pers. Psychol. Comp.* 2, 1667–1684. doi: 10.1111/j.1751-9004.2008.00122.x

Foot, P. (1978). *Virtues and Vices.* Berkeley, CA: University of California Press.

Fowers, B. J. (2012). An aristotelian framework for the human good. *J. Theor. Philos. Psychol.* 32, 10–23. doi: 10.1037/a0025820

Franklin, V. L., Waller, A., Pagliari, C., and Greene, S. A. (2006). A randomized controlled trial of Sweet Talk, a text−messaging system to support young people with diabetes. *Diabet. Med.* 23, 1332–1338. doi: 10.1111/j.1464-5491.2006.01989.x

Freude, G., and Ullsperger, P. (1987). The readiness potential—movement-correlated changes in bioelectric brain activity. *Psychiatrie Neurologie Medizinische Psychologie* 39, 449–459.

Furr, M. (2009). Profile analysis in person–situation integration. *J. Res. Pers.* 43, 196–207. doi: 10.1016/j.jrp.2008.08.002

Gable, S. L., Shelly, L., and Jonathan, H. (2005). What (and why) is positive psychology? *Rev. Gen. Psychol.* 9, 103–110. doi: 10.1037/1089-2680.9.2.103

Gale, E. (2004). The Hawthorne studies-a fable for our times? *Q. J. Med.* 97, 439–449. doi: 10.1093/qjmed/hch070

Garcia-Palacios, A., Herrero, R., Belmonte, M. A., Castilla, D., Guixeres, J., Molinari, G., et al. (2014). Ecological momentary assessment for chronic pain in fibromyalgia using a smartphone: a randomized crossover study. *Eur. J. Pain* 18, 862–872. doi: 10.1002/j.1532-2149.2013.00425.x

Gawronski, B., and Cesario, J. (2013) Of mice and men: what animal research can tell us about context effects on automatic responses in humans. *Pers. Soc. Psychol. Rev.* 17, 187–215. doi: 10.1177/1088868313480096

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8174–8179. doi: 10.1073/pnas.0402680101

Gonzalez, V. M., and Dulin, P. L. (2015). Comparison of a smartphone app for alcohol use disorders with an internet-based intervention plus bibliotherapy: a pilot study. *J. Consult. Clin. Psychol.* 83, 335–345. doi: 10.1037/a0038620

Goodwin, M. S., Velicer, W. F., and Intille, S. S. (2008). Telemetric monitoring in the behavioral sciences. *Behav. Res. Methods* 40, 328–341. doi: 10.3758/BRM.40.1.328

Granholm, E., Ben-Zeev, D., Link, P. C., Bradshaw, K. R., and Holden, J. L. (2012). Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. *Schizophr. Bull.* 38, 414–425. doi: 10.1093/schbul/sbr155

Greenfield, E. A., and Marks, M. F. (2007). Religious social identity as an explanatory factor for associations between more frequent formal religious participation and psychological well-being. *Int. J. Psychol. Relig.* 17, 245–259. doi: 10.1080/10508610701402309

Hammaker, E. L. (2012). "Why researchers should think 'within-person': a paradigmatic rationale," in *Handbook of Research Methods for Studying Daily Life*, eds M. Mehl and T. Conner (New York, NY: Guilford), 43–61.

Hammond, J. S., Keeney, R. L., and Raiffa, H. (1998). The hidden traps in decision making. *Harv. Bus. Rev.* 76, 47–58.

Haney, C., Banks, W. C., and Zimbardo, P. G. (1973). A study of prisoners and guards in a simulated prison. *Naval Res. Rev.* 30, 4–17.

Harman, G. (1999). Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error. *Proc. Aristotelian Soc.* 99, 315–331. doi: 10.1111/1467-9264.00062

Harris, F., and Lahey, B. B. (1982). Subject reactivity in direct observational assessment: a review and critical analysis. *Clin. Psychol. Rev.* 2, 523–538. doi: 10.1016/0272-7358(82)90028-9

Hartshorne, H., and May, M. A. (1928). *Studies in the Nature of Character: Studies in Deceit*. London: Macmillan.

He, F., Zhou, Q., Zhao, Z., Zhang, Y., and Guan, H. (2014). Effects of perceived social support and dispositional optimism on the depression of burn patients. *J. Health Psychol.* doi: 10.1177/1359105314546776. [Epub ahead of print].

Held, B. S. (2005). The "Virtues" of positive psychology. *J. Theor. Philos. Psychol.* 25, 1–34. doi: 10.1037/h0091249

Heron, K. E., and Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *Br. J. Health Psychol.* 15, 1–39. doi: 10.1348/135910709X466063

Hill, P. C., and Pargament, K. I. (2003). Advances in the conceptualization and measurement of religion and spirituality: implications for physical and mental health research. *Am. Psychol.* 58, 64–74. doi: 10.1037/0003-066X.58.1.64

Hofmann, W., Baumeister, R. F., Förster, G., and Vohs, K. D. (2012). Everyday temptations: an experience sampling study of desire, conflict, and self-control. *J. Pers. Soc. Psychol.* 102, 1318–1335. doi: 10.1037/a0026545

Hofmann, W., Wiseneski, D. C., Brandt, M. J., and Skitka, L. J. (2014). Morality in everyday life. *Science* 345, 1340–1343. doi: 10.1126/science.1251560

Hollett, T., and Leander, K. M. (2013). "Location-based environments and technologies," in *The SAGE Handbook of Digital Technology Research*, eds S. Price, C. Jewitt, and B. Brown (London: Sage).

Holmqvist, M., Vincent, N., and Walsh, K. (2014). Web-vs telehealth-based delivery of cognitive behavioral therapy for insomnia: a randomized controlled trial. *Sleep Med.* 15, 187–195. doi: 10.1016/j.sleep.2013.10.013

Hötzel, K., von Brachel, R., Schlossmacher, L., and Vocks, S. (2013). Assessing motivation to change in eating disorders: a systematic review. *J. Eat. Disord.* 1, 38–39. doi: 10.1186/2050-2974-1-38

Huang, J. H., and Luo, G. Y. (2014). Irregular sampling of GPS signal for real-time implementation of pipelined PFFT. *Appl. Mech. Materials* 519, 969–974. doi: 10.4028/www.scientific.net/AMM.519-520.967

Huffziger, S., Ebner-Priemer, U., Eisenbach, C., Koudela, S., Reinhard, I., Zamoscik, V., et al. (2013). Induced ruminative and mindful attention in everyday life: an experimental ambulatory assessment study. *J. Behav. Ther. Exp. Psychiatry* 44, 322–328. doi: 10.1016/j.jbtep.2013.01.007

Hummer, R. A., Rogers, R. G., Nam, C. B., and Ellison, C. G. (1999). Religious involvement and US adult mortality. *Demography* 36, 273–285. doi: 10.2307/2648114

Hunt, J. M. (1965). Traditional personality theory in the light of recent evidence. *Am. Sci.* S3, 80–96.

Hurling, R., Catt, M., De Boni, M., Fairley, B. M., Hurst, T., Murray, P., et al. (2007). Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. *J. Med. Inter. Res.* 9:e7. doi: 10.2196/jmir.9.2.e7

Hyman, J. (2014). Desires, dispositions and deviant causal chains. *Philosophy* 89, 83–112. doi: 10.1017/S0031819113000685

Ibanez, A., Melloni, M., Huepe, D., Helgiu, E., Rivera-Rei, A., Canales-Johnson, A., et al. (2012). What event-related potentials (ERPs) bring to social neuroscience? *Soc. Neurosci.* 7, 632–649. doi: 10.1080/17470919.2012.691078

Isen, A. M., and Levin, P. F. (1972). Effect of feeling good on helping: cookies and kindness. *J. Pers. Soc. Psychol.* 21, 384–388. doi: 10.1037/h0032317

McNulty, J. K., and Fincham, F. D. (2012). Beyond positive psychology? Toward a contextual view of psychological processes and well-being. *Am. Psychol.* 67, 101–210. doi: 10.1037/a0024572

Jankowski, T., and Holas, P. (2014). Metacognitive model of mindfulness. *Conscious. Cogn.* 28, 64–80. doi: 10.1016/j.concog.2014.06.005

Jayawickreme, E., and Chemero, A. (2008). Ecological moral realism: an alternative theoretical framework for studying moral psychology. *Rev. Gen. Psychol.* 12, 118–126. doi: 10.1037/1089-2680.12.2.118

Jha, A. P., Stanley, E. A., Kiyonaga, A., Wong, L., and Gelfand, L. (2010). Examining the protective effects of mindfulness training on working memory capacity and affective experience. *Emotion* 10, 54–64. doi: 10.1037/a0018438

John, O. P., Naumann, L. P., and Soto, C. J. (2008). "Paradigm shift to the integrative Big Five trait taxonomy," in *Handbook of Personality: Theory and Research*, 3rd Edn., eds O. P. John, R. W. Robbins, and L. A. Pervin (New York, NY: Guilford), 114–158.

Joo, N. S., and Kim, B. T. (2007). Mobile phone short message service messaging for behaviour modification in a community-based weight control programme in Korea. *J. Telemed. Telecare* 13, 416–420. doi: 10.1258/135763307783064331

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: past, present, and future. *Clin. Psychol. Sci. Pract.* 10, 144–156. doi: 10.1093/clipsy.bpg016

Kaltenthaler, E., Brazier, J., De Nigris, E., Tumur, I., Ferriter, M., Beverley, C., et al. (2006). Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. *Health Technol. Assess.* 10, 1–186. doi: 10.3310/hta10330

Kamin, L. J. (1968). "'Attention-like' processes in classical conditioning," in *Miami Symposium on the Prediction of Behavior, 1967: Aversive Stimulation*, ed M. R. Jones (Coral Gables, FL: University of Miami Press), 9–31.

Kane, M. J., Brown, L. H., McVay, J. C., Myin-Germeys, I., and Kwapil, T. R. (2007). For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. *Psychol. Sci.* 18, 614–621. doi: 10.1111/j.1467-9280.2007.01948.x

Kauer, S. D., Reid, S. C., Crooke, A. H. D., Khor, A., Hearps, S. J. C., Jorm, A. F., et al. (2012). Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. *J. Med. Inter. Res.* 14, e67. doi: 10.2196/jmir.1858

Keedwell, P. A., and Linden, D. E. J. (2013). Integrative neuroimaging in mood disorders. *Curr. Opin. Psychiatry* 26, 27–32. doi: 10.1097/YCO.0b013e32835a0b63

Kelley, A. E., Schochet, T., and Landry, C. F. (2004). Risk taking and novelty seeking in adolescence: introduction to part I. *Ann. N.Y. Acad. Sci.* 1021, 27–32. doi: 10.1196/annals.1308.003

Keltner, D. (2009). *Born to be Good: The Science of a Meaningful Life*. New York, NY: WW Norton & Company.

Kenny, A. (2011). "Introduction," in *The Eudemian Ethics, Aristotle*, trans ed A. Kenny (Oxford: Oxford University Press).

Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. *Am. Psychol.* 62, 95–108. doi: 10.1037/0003-066X.62.2.95

Khor, A. S., Gray, K. M., Reid, S. C., and Melvin, G. A. (2014). Feasibility and validity of ecological momentary assessment in adolescents with high-functioning autism and Asperger's disorder. *J. Adoles.* 37, 37–46. doi: 10.1016/j.adolescence.2013.10.005

Kim, H. S., and Jeong, H. S. (2007). A nurse short message service by cellular phone in type−2 diabetic patients for six months. *J. Clin. Nurs.* 16, 1082–1087. doi: 10.1111/j.1365-2702.2007.01698.x

King, A. C., Ahn, D. K., Oliveira, B. M., Atienza, A. A., Castro, C. M., and Gardner, C. D. (2008). Promoting physical activity through hand-held computer technology. *Am. J. Prevent. Med.* 34, 138–142. doi: 10.1016/j.amepre.2007.09.025

King, A. C., Heckler, E. B., Grieco, L. A., Winter, S. J., Sheats, J. L., Buman, M. P., et al. (2013). Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. *PLoS ONE* 8:e62613. doi: 10.1371/journal.pone.0062613

Kirk, G. D., Linas, B. S., Westergaard, R. P., Piggott, D., Bollinger, R. C., Chang, L. W., et al. (2013). The exposure assessment in current time study: implementation, feasibility, and acceptability of real-time data collection in a community cohort of illicit drug users. *AIDS Res. Treat.* 2013:594671. doi: 10.1155/2013/594671

Klimecki, O. M., Leiberg, S., Lamm, C., and Singer, T. (2012). Functional neural plasticity and associated changes in positive affect after compassion training. *Cereb. Cortex* 9, 873–879. doi: 10.1093/cercor/bhs142

Klimecki, O. M., Leiberg, S., Ricard, M., and Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Soc. Cogn. Affect. Neurosci.* 9, 873–879. doi: 10.1093/scan/nst060

Klingberg, T. (2010). Training and plasticity of working memory. *Trends Cogn. Sci.* 14, 317–324. doi: 10.1016/j.tics.2010.05.002

Knaevelsrud, C., and Maercker, A. (2007). Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial. *BMC Psychiatry* 7:13. doi: 10.1186/1471-244X-7-13

Konrath, S. (2015). "Positive technology. Using mobile phones for psychosocial interventions," in *Encyclopedia of Mobile Phone Behavior*, ed Z. Yan (Hershey, PA: IGI Global).

Korotitsch, W. J., and Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychol. Assess.* 11, 415–425. doi: 10.1037/1040-3590.11.4.415

Koval, P., Ogrinz, B., Kuppens, P., Van den Bergh, O., Tuerlinckx, F., and Sütterlin, S. (2013). Affective instability in daily life is predicted by resting heart rate variability. *PLoS ONE* 8:e81536. doi: 10.1371/journal.pone.0081536

Koven, N. S., and Max, L. K. (2014). Basal salivary oxytocin level predicts extra-but not intra-personal dimensions of emotional intelligence. *Psychoneuroendocrinology* 44, 20–29. doi: 10.1016/j.psyneuen.2014.02.018

Kramer, I., Simons, C. J., Hartmann, J. A., Mennet-Lothmann, C., Viechtbauer, W., and Peelers, F. (2014). A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry* 13, 68–77. doi: 10.1002/wps.20090

Kreuter, M., Farrell, D., Olevitch, L., and Brennan, L. (2000a). *Tailoring Health Messages: Customizing Communication with Computer Technology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kreuter, M. W., Oswald, D. L., Bull, F. C., and Clark, E. M. (2000b). Are tailored health education materials always more effective than non-tailored materials? *Health Educ. Res.* 15, 305–315. doi: 10.1093/her/15.3.305

Kristjánsson, K. (2010). Positive psychology, happiness, and virtue: the troublesome conceptual issues. *Rev. Gen. Psychol.* 14, 296–310. doi: 10.1037/a0020781

Kristjánsson, K. (2013). *Virtues and Vices in Positive Psychology*. Cambridge: Cambridge University Press.

Kuo, B. C. H., Arnold, R., and Rodriguez-Rubio, B. (2014). Mediating effects of coping in the link between spirituality and psychological distress in a culturally diverse undergraduate sample. *Mental Health Relig. Cult.* 17, 173–184. doi: 10.1080/13674676.2013.780015

Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502. doi: 10.1016/j.neuroimage.2010.10.014

Lane, N. D., Lin, M., Mohammod, M., Yang, X., Lu, H., Cardone, G., et al. (2014). BeWell: sensing sleep, physical activities and social interactions to promote wellbeing. *Mob. Netw. Appl.* 19, 345–359. doi: 10.1007/s11036-013-0484-5

Lappalainen, P., Granlund, A., Siltanen, S., Ahonen, S., Vitikainen, M., Tolvanen, A., et al. (2014). ACT Internet-based vs face-to-face? A randomized controlled trial of two ways to deliver acceptance and commitment therapy for depressive symptoms: an 18-month follow-up. *Behav. Res. Ther.* 61, 43–54. doi: 10.1016/j.brat.2014.07.006

Lawler-Row, K. A. (2010). Forgiveness as a mediator of the religiosity—Health relationship. *Psychol. Relig. Spirit.* 2, 1–16. doi: 10.1037/a0017584

Levesque, C., and Brown, K. W. (2007). Mindfulness as a moderator of the effect of implicit motivational self-concept on day-to-day behavioral motivation. *Motivat. Emot.* 31, 284–299. doi: 10.1007/s11031-007-9075-8

Levey, A. B., and Martin, I. (1975). Classical conditioning of human 'evaluative' responses. *Behav. Res. Ther.* 13, 221–226. doi: 10.1016/0005-7967(75)90026-1

Linden, D. E. J. (2014). Neurofeedback and networks of depression. *Dialog. Clin. Neurosci.* 16, 103–112.

Litz, B. T., Engel, C. C., Bryant, R. A., and Papa, A. (2007). A randomized, controlled proof-of-concept trial of an Internet-based, therapist-assisted self-management treatment for posttraumatic stress disorder. *Am. J. Psychiatry* 164, 1676–1683. doi: 10.1176/appi.ajp.2007.06122057

Lopez, R., Hofmann, W., Wagner, D. D., Kelly, W. M., and Heatherton, T. F. (2014). Neural predictors of giving in to temptation in daily life. *Psychol. Sci.* 25, 1337–1344. doi: 10.1177/0956797614531492

Luszczynska, A., Gutierrez-Dona, B., and Schwarzer, R. (2005). General self-efficacy in various domains of human functioning: evidence from five countries. *Int. J. Psychol.* 40, 80–89. doi: 10.1080/00207590444000041

Lutz, A., Slagter, H. A., Dunne, J. D., and Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends Cogn. Sci.* 12, 163–169. doi: 10.1016/j.tics.2008.01.005

Maas, J., Hietbrink, L., Rinck, M., and Keijsers, G. P. J. (2013). Changing automatic behavior through self-monitoring: does overt change also imply implicit change? *J. Behav. Ther. Exp. Psychiatry* 44, 279–284. doi: 10.1016/j.jbtep.2012.12.002

MacBeth, A., and Gumley, A. (2012). Exploring compassion: a meta-analysis of the association between self-compassion and psychopathology. *Clin. Psychol. Rev.* 32, 545–552. doi: 10.1016/j.cpr.2012.06.003

Macias, C., Panch, T., Hicks, Y. M., Scolnick, J. S., Weene, D. L., Öngür, D., et al. (2015). Using smartphone apps to promote psychiatric and physical well-being. *Psychiatric Q.* doi: 10.1007/s11126-015-9337-7. [Epub ahead of print].

MacIntyre, A. (1984). *After Virtue*. Notre Dame, IN: University of Notre Dame Press.

MacIntyre, A. (1999). *Dependent Rational Animals: why Human Beings Need the Virtues*. Chicago, IL: Open Court.

MacKerron, G., and Mourato, S. (2013). Happiness is greater in natural environments. *Glob. Environ. Change* 23, 992–1000. doi: 10.1016/j.gloenvcha.2013.03.010

MacLean, K. A., Ferrer, E., Aichele, S. R., Bridwell, D. A., Zanesco, A. P., Jacobs, T. A., et al. (2010). Intensive meditation training improves perceptual discrimination and sustained attention. *Psychol. Sci.* 21, 829–839. doi: 10.1177/0956797610371339

Maddison, R., Pfaeffli, L., Stewart, R., Kerr, A., Jiang, Y., Rawstorn, J., et al. (2014). The HEART mobile phone trial: the partial mediating effects of self-efficacy on physical activity among cardiac patients. *Front. Public Health* 2:56. doi: 10.3389/fpubh.2014.00056

Magidson, J. F., Roberts, B. W., Collado-Rodriguez, A., and Lejuez, C. W. (2014). Theory-driven intervention for changing personality: expectancy value theory, behavioral activation, and conscientiousness. *Dev. Psychol.* 50, 1442–1450. doi: 10.1037/a0030583

Maltby, J., and Day, L. (2001). Spiritual involvement and belief: the relationship between spirituality and Eysenck's personality dimensions. *Pers. Indiv. Diff.* 30, 187–192. doi: 10.1016/S0191-8869(00)00024-6

March, S., Spence, S. H., and Donovan, C. L. (2009). The effects of an internet-based cognitive-behavioral therapy intervention for child anxiety disorders. *J. Pediatric Psychol.* 34, 474–487. doi: 10.1093/jpepsy/jsn099

Marchant, N. J., Hamlin, A. S., and McNally, G. P. (2009). Lateral hypothalamus is required for context-induced reinstatement of extinguished reward seeking. *J. Neurosci.* 29, 1331–1342. doi: 10.1523/JNEUROSCI.5194-08.2009

Marchant, N. J., Rabei, R., Kaganovsky, K., Caprioli, D., Bossert, J. M., Bonci, A., et al. (2014). A critical role of lateral hypothalamus in context-induced relapse to alcohol seeking after punishment-imposed abstinence. *J. Neurosci.* 34, 7447–7457. doi: 10.1523/JNEUROSCI.0256-14.2014

Marco, C. A., and Suls, J. (1993). Daily stress and the trajectory of mood: spillover, response assimilation, contrast and chronic negative affectivity. *J. Pers. Soc. Psychol.* 64, 1053–1063. doi: 10.1037/0022-3514.64.6.1053

Marsh, C. (2014). *Strange Glory: A Life of Dietrich Bonhoeffer*. New York, NY: Knopf.

McCambridge, J., Witton, J., and Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J. Clin. Epidemiol.* 67, 267–277. doi: 10.1016/j.jclinepi.2013.08.015

McCrae, R. R., and Costa, P. T. (2003). *Personality in Adulthood: a Five-Factor Theory Perspective*. New York, NY: The Guilford Press.

McCullough, M. E., Hoyt, W. T., Larson, D. B., Koenig, H. G., and Thoresen, C. (2000). Religious involvement and mortality: a meta-analytic review. *Health Psychol.* 19, 211–222. doi: 10.1037/0278-6133.19.3.211

McNiel, J. M., and Fleeson, W. (2006). The causal effects of extraversion on positive affect and neuroticism on negative affect: manipulating state extraversion and state neuroticism in an experimental approach. *J. Res. Pers.* 40, 529–550. doi: 10.1016/j.jrp.2005.05.003

Mehl, M. R., Robbins, M. L., and Deters, F. G. (2012). Naturalistic observation of health-relevant social processes: the electronically activated recorder methodology in psychosomatics. *Psychosom. Med.* 74, 410–417. doi: 10.1097/PSY.0b013e3182545470

Merchant, H., Harrington, D. L., and Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annu. Rev. Neurosci.* 36, 313–336. doi: 10.1146/annurev-neuro-062012-170349

Mhurchu, C. N., Whittaker, R., McRobbie, H., Ball, K., Crawford, D., Michie, J., et al. (2014). Feasibility, acceptability and potential effectiveness of a mobile health (mHealth) weight management programme for New Zealand adults. *BMC Obesity* 1:10. doi: 10.1186/2052-9538-1-10

Milgram, S. (1963). Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* 67, 371–378. doi: 10.1037/h0040525

Miller, C. B. (2013). *Moral Character: An Empirical Theory*. Oxford: Oxford University Press.

Miller, W. R., and Thoresen, C. E. (2003). Spirituality, religion, and health: an emerging research field. *Am. Psychol.* 58, 24–35. doi: 10.1037/0003-066X.58.1.24

Minbashian, A., Wood, R. E., and Beckmann, N. (2010). Task-contingent conscientiousness as a unit of personality at work. *J. Appl. Psychol.* 95, 793–806. doi: 10.1037/a0020016

Mischel, W. (1968). *Personality and Assessment*. New York, NY: Wiley.

Moberly, N. J., and Watkins, E. R. (2008). Ruminative self-focus and negative affect: an experience sampling study. *J. Abnorm. Psychol.* 117, 314–323. doi: 10.1037/0021-843X.117.2.314

Moisa, M., Siebner, H. R., Pohmann, R., and Thielscher, A. (2012). Uncovering a context-specific connectional fingerprint of human dorsal premotor cortex. *J. Neurosci.* 32, 7244–7252. doi: 10.1523/JNEUROSCI.2757-11.2012

Moore, B. A., Fazzino, T., Garnet, B., Cutter, C. J., and Barry, D. T. (2011). Computer-based interventions for drug use disorders: a systematic review. *J. Subst. Abus. Treat.* 40, 215–223. doi: 10.1016/j.jsat.2010.11.002

Morris, M. E., Kathwala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., Labhard, M., et al. (2010). Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. *J. Med. Inter. Res.* 12, e10. doi: 10.2196/jmir.1371

Moskowitz, D. S., Pinard, G., Zuroff, D. C., Annable, L., Stat, D., and Young, S. N. (2001). The effect of tryptophan on social interaction in every day life: a placebo-controlled study. *Neuropsychopharmacology* 25, 277–289. doi: 10.1016/S0893-133X(01)00219-6

Moskowitz, D. S., and Young, S. N. (2006). Ecological momentary assessment: What it is and why it is a method of the future in clinical psychopharmacology. *J. Psychiatry Neurosci.* 31, 13–20. doi: 10.1186/1471-2369-15-29

Mumford, S. (1998). *Dispositions*. Oxford: Oxford University Press.

Musiat, P., and Tarrier, N. (2014). Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. *Psychol. Med.* 44, 1–14. doi: 10.1017/S0033291714000245

Nachev, P., and Hacker, P. (2014). The neural antecedents to voluntary action: a conceptual analysis. *Cogn. Psychol.* 5, 193–218. doi: 10.1080/17588928.2014.934215

Nahmias, E. (2010). "Introduction," in *Moral Psychology: Historical and Contemporary Readings*, eds T. Nadelhoffer, E. Nahmias, and S. Nichols (Oxford: Wiley-Blackwell), 163–166.

Neal, D. T., Wood, W., and Quinn, J. M. (2006). Habits—a repeat performance. *Curr. Dir. Psychol. Sci.* 15, 196–202. doi: 10.1111/j.1467-8721.2006.00435.x

Neal, D. T., Wood, W., Wu, M., and Kurlander, D. (2011). The pull of the past: when do habits persist despite conflict with motives? *Pers. Soc. Psychol. Bull.* 37, 1428–1437. doi: 10.1177/0146167211419863

Newman, M. G., Szkodny, L. E., Llera, S. J., and Przeworski, A. (2011). A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? *Clin. Psychol. Rev.* 31, 89–103. doi: 10.1016/j.cpr.2010.09.008

Nix, G. A., Ryan, R. M., Manly, J. B., and Deci, E. L. (1999). Revitalization through self-regulation: the effects of autonomous and controlled motivation on happiness and vitality. *J. Exp. Soc. Psychol.* 35, 266–284. doi: 10.1006/jesp.1999.1382

O'Rourke, N. (2004). Psychological resilience and the well-being of widowed women. *Ageing Int.* 29, 267–280. doi: 10.1007/s12126-996-1002-x

Obermayer, J. L., Riley, W. T., Asif, O., and Jean-Mary, J. (2004). College smoking-cessation using cell phone text messaging. *J. Am. Coll. Health* 53, 71–78. doi: 10.3200/JACH.53.2.71-78

Oman, D., and Thoresen, C. E. (2005). "Do religion and spirituality influence health," in *Handbook of the Psychology of Religion and Spirituality*, eds R. F. Paloutzian and C. L. Park (New York, NY: Guilford Press), 435–459.

Ong, A. D., Bergeman, C. S., Bisconti, T. L., and Wallace, K. A. (2006). Psychological resilience, positive emotions, and successful adaptation to stress in later life. *J. Pers. Soc. Psychol.* 91, 730–749. doi: 10.1037/0022-3514.91.4.730

Ortner, C. N. M., Kilner, S. J., and Zelazo, P. D. (2007). Mindfulness meditation and reduced emotional interference on a cognitive task. *Motiv. Emot.* 31, 271–283. doi: 10.1007/s11031-007-9076-7

Ottaviani, C., Medea, B., Lonigro, A., Tarvainen, M., and Couyoumdjian, A. (2015). Cognitive rigidity is mirrored by autonomic inflexibility in daily life perseverative cognition. *Biol. Psychol.* 107, 24–30. doi: 10.1016/j.biopsycho.2015.02.011

Ozer, D. J. (1986). *Consistency in Personality: A Methodological Framework*. Berlin: Springer-Verlag.

Palmier-Claus, J. E., Ainsworth, J., Machin, M., Barrowclough, C., Dunn, G., Barkus, E., et al. (2012). The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. *BMC Psychiatry* 12:172. doi: 10.1186/1471-244X-12-172

Park, C. L. (2007). Religiousness/spirituality and health: a meaning systems perspective. *J. Behav. Med.* 30, 319–328. doi: 10.1007/s10865-007-9111-x

Park, N., Peterson, C., and Seligman, M. (2006). Character strengths in fifty-four nations and the fifty US states. *J. Posit. Psychol.* 1, 118–129. doi: 10.1080/17439760600619567

Parsons, H. (1974). What happened at Hawthorne? *Science* 183, 922–932. doi: 10.1126/science.183.4128.922

Patrick, K., Raab, F., Adams, M. A., Dillon, L., Zabinski, M., Rock, C. L., et al. (2009). A text message-based intervention for weight loss: randomized controlled trial. *J. Med. Inter. Res.* 11, e1. doi: 10.2196/jmir.1100

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *J. Learn. Sci.* 13, 423–451. doi: 10.1207/s15327809jls1303_6

Peck, C. A., and Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learn. Motiv.* 21, 1–31. doi: 10.1016/0023-9690(90)90002-6

Peterson, C., and Seligman, M. (2004). *Character Strengths and Virtues: a Handbook and Classification*. Washington, DC: American Psychological Association.

Pfefferbaum, A., Mathalon, D. H., Sullivan, E. V., Rawles, J. M., Zipursky, R. B., and Lim, K. O. (1994). A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. *JAMA Neurol.* 51, 874–887. doi: 10.1001/archneur.1994.00540210046012

Plato, (380 BC/1991). *Republic*. Transl. A. Bloom. New York, NY: Basic Books.

Plaza, I., Demarzo, M. M. P., Herrera-Mercadal, P., and García-Campayo, J. (2013). Mindfulness-based mobile applications: literature review and analysis of current features. *J. Med. Inter. Res.* 1, e24. doi: 10.2196/mhealth.2733

Pop-Eleches, C., Thirumurthy, H., Habyarimana, J. P., Zivin, J. G., Goldstein, M. P., De Walque, D., et al. (2011). Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized

controlled trial of text message reminders. *AIDS (Lond.)* 25, 825–834. doi: 10.1097/QAD.0b013e32834380c1

Praamstra, P., Kourtis, D., and Nazarpour, K. (2009). Simultaneous preparation of multiple potential movements: opposing effects of spatial proximity mediated by premotor and parietal cortex. *J. Neurophysiol.* 102, 2084–2095. doi: 10.1152/jn.00413.2009

Pramana, G., Parmanto, B., Kendall, P. C., and Silk, J. S. (2014). The SmartCAT: an m-Health platform for ecological momentary intervention in child anxiety treatment. *Telemed. e-Health* 20, 419–427. doi: 10.1089/tmj.2013.0214

Przeworski, A., and Newman, M. G. (2004). Palmtop computer-assisted group therapy for social phobia. *J. Clin. Psychol.* 60, 179–188. doi: 10.1002/jclp. 10246

Quinn, J. M., Pascoe, A., Wood, W., and Neal, D. T. (2010). Can't control yourself? Monitor those bad habits. *Pers. Soc. Psychol. Bull.* 36, 499–511. doi: 10.1177/0146167209360665

Raento, M., Oulasvirta, A., and Eagle, N. (2009). Smartphones an emerging tool for social scientists. *Soc. Methods Res.* 37, 426–454. doi: 10.1177/0049124108330005

Räikkönen, K., Matthew, K. A., Flory, J. D., Owens, J. F., and Gump, B. B. (1999). Effects of optimism, pessimism, and trait anxiety on ambulatory blood pressure and mood during everyday life. *J. Pers. Soc. Psychol.* 76, 104–113. doi: 10.1037/0022-3514.76.1.104

Randall, W. M., Rickard, N. S., and Vella-Brodrick, D. A. (2014). Emotional outcomes of regulation strategies used during personal music listening: a mobile experience sampling study. *Musicae Scientiae* 18, 275–291. doi: 10.1177/1029864914536430

Reiss, S., and Havercamp, S. (1996). The sensitivity theory of motivation: implications for psychopathology. *Behav. Res. Ther.* 34, 621–632. doi: 10.1016/0005-7967(96)00041-1

Rescorla, R. A., and Solomon, R. L. (1967). Two-process learning theory: relationships between pavlovian conditioning and instrumental learning. *Psychol. Rev.* 74, 151–182. doi: 10.1037/h0024475

Rescorla, R. A., and Wagner, A. R. A. (1972). "A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement," in *Classical Conditioning II: Current Research and Theory*, eds A. Black and W. F. Prokasy (New York, NY: Appelton-Century-Crofts), 64–99.

Reutter, K. K., and Bigatti, S. M. (2014). Religiosity and spirituality as resiliency resources: moderation, mediation, or moderated mediation? *J. Sci. Study Relig.* 53, 56–72. doi: 10.1111/jssr.12081

Rizvi, S. L., Dimeff, L. A., Skutch, J., Carroll, D., and Linehan, M. M. (2011). A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. *Behav. Ther.* 42, 589–600. doi: 10.1016/j.beth.2011.01.003

Robinson, E., Aveyard, P., Daley, A., Jolly, K., Lewis, A., Lycett, D., et al. (2013). Eating attentively: a systematic review and meta-analysis of the effect of food intake memory and awareness on eating. *Am. J. Clin. Nutr.* 97, 728–742. doi: 10.3945/ajcn.112.045245

Rodgers, A., Corbett, T., Bramley, D., Riddell, T., Wills, M., Lin, R. B., et al. (2005). Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. *Tobacco Control* 14, 255–261. doi: 10.1136/tc.2005.011577

Roesch, S. C., Aldridge, A. A., Stocking, S. N., Villodas, F., Leung, Q., Bartley, C. E., et al. (2010). Multilevel factor analysis and structural equation modeling of daily diary coping data: modeling trait and state variation. *Multiv. Behav. Res.* 45, 767–789. doi: 10.1080/00273171.2010.519276

Ross, L., and Nisbett, R. E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. New York, NY: Mcgraw-Hill Book Company.

Rossi, N. E., Bisconti, T. L., and Bergeman, C. S. (2007). The role of dispositional resilience in regaining life satisfaction after the loss of a spouse. *Death Stud.* 31, 863–883. doi: 10.1080/07481180701603246

Runyan, J. D., Moore, A. N., and Dash, P. (2004). A role for the prefrontal cortex in memory storage for trace fear conditioning. *J. Neurosci.* 24, 1288–1295. doi: 10.1523/JNEUROSCI.4880-03.2004

Runyan, J. D., Steenbergh, T. A., Bainbridge, C., Daugherty, D. A., Oke, L., and Fry, B. N. (2013). A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness. *PLoS ONE* 8:e71325. doi: 10.1371/journal.pone.0071325

Rushton, P. (1980). *Altruism, Socialization, and Society*. Englewood Cliffs, NJ: Prentice-Hall.

Rushton, P. (1984). The altruistic personality. *Dev. Mainten. Prosoc. Behav.* 31, 271–290.

Rushton, P., Brainerd, C. J., and Pressley, M. (1983). Behavioral development and construct validity: the principle of aggregation. *Psychol. Bull.* 94, 18–38. doi: 10.1037/0033-2909.94.1.18

Rushton, P., Chrisjohn, R. D., and Fekken, C. G. (1981). The altruistic personality and the self-report altruism scale. *Pers. Indiv. Diff.* 2, 293–302. doi: 10.1016/0191-8869(81)90084-2

Ruwaard, J., Broeksteeg, J., Schrieken, B., Emmelkamp, P., and Lange, A. (2010). Web-based therapist-assisted cognitive behavioral treatment of panic symptoms: a randomized controlled trial with a three-year follow-up. *J. Anx. Disord.* 24, 387–396. doi: 10.1016/j.janxdis.2010.01.010

Ryan, R. M., and Deci, E. L. (2001). On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. *Annu. Rev. Psychol.* 52, 141–166. doi: 10.1146/annurev.psych.52.1.141

Rydell, R. J., and Gawronski, B. (2009). I like you, I like you not: understanding the formation of context-dependent automatic attitudes. *Cogn. Emot.* 23, 1118–1152. doi: 10.1080/02699930802355255

Ryff, C. D. (2013). "Eudaimonic well-being and health: mapping consequences of self-realization," in *The Best Within Us: Positive Psychology Perspectives on Eudaimonia*, ed A. S. Waterman (Washington, DC: American Psychological Association), 77–89.

Ryff, C. D., and Singer, B. H. (2008). Know thyself and become what you are: a eudaimonic approach to psychological well-being. *J. Happin. Stud.* 9, 13–39. doi: 10.1007/s10902-006-9019-0

Salomon, G. (1993). "On the nature of pedagogic computer tools: the case for the writing partner," in *Computers as Cognitive Tools*, eds S. P. Lajoie and S. J. Derry (Mahwah, NJ: Lawrence Erlbaum Associates), 179–196.

Schachtman, T. R., and Reilly, S. S. (2011). "Things you always wanted to know about learning theory but were afraid to ask," in *Associative Learning and Conditioning Theory: Human and Non-human Applications*, eds T. R. Schachtman and S. S. Reilly (New York, NY: Oxford University Press), 3–26.

Scheier, M. E., and Carver, C. S. (1987). Dispositional optimism and physical well-being: the influence of generalized outcome expectancies on health. *J. Pers.* 55, 169–210. doi: 10.1111/j.1467-6494.1987.tb00434.x

Scheier, M. F., Carver, C. S., and Bridges, K. (2001). "Optimism, pessimism, and psychological well-being," in *Optimism and Pessimism: Implications for Theory, Research, and Practice*, ed E. C. Change (Washington, DC: American Psychological Association), 127–145.

Scheier, M. F., Matthews, K. A., Owens, J. F., Magovern, G. F., Lefebvre, R. C., Abbott, R. A., et al. (1989). Dispositional optimism and recovery from coronary artery bypass surgery: the beneficial effects on physical and psychological well-being. *J. Pers.Soc. Psychol.* 57, 1024–1040. doi: 10.1037/0022-3514.57.6.1024

Schlicht, W., Ebner-Priemer, U. W., and Kanning, M. (2013). Ecological momentary assessment and intervention in physical activity and well-being: affective reactions, social-cognitive factors, and behaviors as determinants of physical activity and exercise. *Front. Psychol.* 4:916. doi: 10.3389/fpsyg.2013.00916

Schmajuk, N. A., and Holland, P. C. (1998). *Occasion Setting: Associative Learning and Cognition in Animals*. Washington, DC: American Psychological Association.

Schoenberg, P. L. A., and David, A. S. (2014). Biofeedback for psychiatric disorders: a systematic review. *Appl. Psychophysiol. Biofeedback* 39, 109–135. doi: 10.1007/s10484-014-9246-9

Schwartz, J. E., Neale, J., Marco, C., Shiffman, S., and Stone, A. A. (1999). Does trait coping exist? A momentary assessment approach to the evaluation of traits. *J. Pers. Soc. Psychol.* 77, 360–369. doi: 10.1037/0022-3514. 77.2.360

Schwarz, N. (2007). "Retrospective and concurrent self-reports: the rationale for real-time data capture," in *The Science of Real-time Data Capture: Self-reports in Health Research*, eds A. A. Stone, S. Schiffman, A. A. Atienza, and L. Nebeling (New York, NY: Oxford University Press), 3–23.

Schwerdtfeger, A. R., and Scheel, S.-M. (2012). Self-esteem fluctuations and cardiac vagal control in everyday life. *Int. J. Psychophysiol.* 83, 328–335. doi: 10.1016/j.ijpsycho.2011.11.016

Segalowitz, S. J., and Davies, P. L. (2004). Charting the maturation of the frontal lobe: an electrophysiological strategy. *Brain Cogn.* 55, 116–133. doi: 10.1016/S0278-2626(03)00283-5

Seligman, M. E. P., and Csikszentmihalyi, M. (2000). Positive psychology: an introduction. *Am. Psychol.* 55, 5–14. doi: 10.1037/0003-066X.55.1.5

Seligman, M. E. P., Steen, T. A., Park, N., and Peterson, C. (2005). Positive psychology progress: empirical validation of interventions. *Am. Psychol.* 60, 410–421. doi: 10.1037/0003-066X.60.5.410

Seybold, K. S., and Hill, P. C. (2001). The role of religion and spirituality in mental and physical health. *Cur. Dir. Psychol. Sci.* 10, 21–24. doi: 10.1111/1467-8721.00106

Shanks, D. A., and Dickinson, A. (1991). Instrumental judgment and performance under variations in action-outcome contingency and contiguity. *Mem. and Cogn.* 19, 353–360. doi: 10.3758/BF03197139

Shannon, B. J., Raichle, M. E., Snyder, A. Z., Fair, D. A., Mills, K. L., Zhang, D., et al. (2011). Premotor functional connectivity predicts impulsivity in juvenile offenders. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11241–11245. doi: 10.1073/pnas.1108241108

Shapiro, E. S., and Cole, C. L. (1999). Self-monitoring in assessing children's problems. *Psychol. Asses.* 11, 448–457. doi: 10.1037/1040-3590.11.4.448

Shapiro, S. L., Carlson, L. E., Astin, J. A., and Freedman, B. (2006). Mechanisms of mindfulness. *J. Clin. Psychol.* 62, 373–386. doi: 10.1002/jclp.20237

Shibasaki, H., and Hallett, M. (2006). What is the Bereitschaftspotential? *Clin. Neurophysiol.* 117, 2341–2356. doi: 10.1016/j.clinph.2006.04.025

Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annual Rev. Clin. Psychol.* 4, 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Shryack, J., Steger, M. F., Krueger, R. F., and Kallie, C. S. (2010). The structure of virtue: an empirical investigation of the dimensionality of the virtues in action inventory of strengths. *Pers. Individ. Differ.* 48, 714–719. doi: 10.1016/j.paid.2010.01.007

Silvia, P. J., Beaty, R. E., Nusbaum, E. C., Eddington, K. M., Levin-Aspenson, H., and Kwapil, T. R. (2014). Everyday creativity in daily life: an experience-sampling study of "little c" creativity. *Psychol. Aesthet. Creativity Arts* 8, 183–188. doi: 10.1037/a0035722

Simon, C. R., and Durand-Bush, N. (2014). Does self-regulation capacity predict psychological well-being in physicians?. *Psychol. Health Med.* 1–11. doi: 10.1080/13548506.2014.936887

Solomon, R. L. (1949). An extension of control group design. *Psychol. Bullet.* 46, 137–150. doi: 10.1037/h0062958

Somerville, L. H., Hare, T., and Casey, B. J. (2011). Frontostriatal maturation predicts cognitive control failure to appetitive cues in adolescents. *J. Cogn. Neurosci.* 23, 2123–2134. doi: 10.1162/jocn.2010.21572

Somerville, L. H., Jones, R. M., and Casey, B. J. (2010). A time of change: behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain Cogn.* 72, 124–133. doi: 10.1016/j.bandc.2009.07.003

Sommer, R. (1968). The Hawthorne dogma. *Psychol. Bullet.* 70, 592–595. doi: 10.1037/h0026728

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., and Toga, A. W. (2003). Mapping cortical change across the human life span. *Nat. Neurosci.* 6, 309–315. doi: 10.1038/nn1008

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., and Toga, A. W. (1999). *In vivo* evidence for post-adolescent brain maturation in frontal and striatal regions. *Nat. Neurosci.* 2, 859–861. doi: 10.1038/13154

Spence, J., Titov, N., Dear, B. F., Johnston, L., Solley, K., Lorian, C., et al. (2011). Randomized controlled trial of Internet-delivered cognitive behavioral therapy for posttraumatic stress disorder. *Depress. Anxiety* 28, 541–550. doi: 10.1002/da.20835

Staub, E. (1974). Helping a distressed person: social, personality, and stimulus determinants. *Adv. Exp. Soc. Psychol.* 7, 293–341. doi: 10.1016/S0065-2601(08)60040-4

Stone, A. A., Schwartz, J. E., Neale, J. E., Shiffman, S., Marco, C. A., Hickcox, M., et al. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *J. Person. Soc. Psychol.* 74, 1670–1680. doi: 10.1037/0022-3514.74.6.1670

Stone, A. A., and Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Ann. Behav. Med.* 16, 199–202.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., and Hufford, M. R. (2002). Patient non-compliance with paper diaries. *BMJ* 324, 1193–1194. doi: 10.1136/bmj.324.7347.1193

Strang, S., Utikal, V., Fischbacher, U., Weber, B., and Falk, A. (2014). Neural correlates of receiving an apology and active forgiveness: an fMRI study. *PLoS ONE* 9:e87654. doi: 10.1371/journal.pone.0087654

Strecher, V. J., Shiffman, S., and West, R. (2005). Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction* 100, 682–688. doi: 10.1111/j.1360-0443.2005.01093.x

Suh, E., Diener, E., and Fujita, F. (1996). Events and subjective well-being: only recent events matter. *J. Person. Soc. Psychol.* 70, 1091–1102. doi: 10.1037/0022-3514.70.5.1091

Sundararajan, L. (2005). Happiness donut: a confucian critique of positive psychology. *J. Theor. Philos. Psychol.* 25, 35–60. doi: 10.1037/h0091250

Sutarto, A. P., Wahab, M. N. A., and Zin, N. M. (2010). Heart rate variability (HRV) biofeedback: a new training approach for operator's performance enhancement. *J. Ind. Eng. Manag.* 3, 176–198. doi: 10.3926/jiem.2010.v3n1.p176-198

Sutherland, N. S., and Mackintoch, N. J. (1971). *Mechanisms of Animal Discrimination Learning.* London: Academic Press.

Tamagawa, R., Giese-Davis, J., Speca, M., Doll, R., Stephen, J., and Carlson, L. E. (2013). Trait mindfulness, repression, suppression, and self−reported mood and stress symptoms among women with breast cancer. *J. Clin. Psychol.* 69, 264–277. doi: 10.1002/jclp.21939

Tang, Y. Y., Ma, Y., Wang, J., Fan, Y., Feng, S., Lu, Q., et al. (2007). Short-term meditation training improves attention and self-regulation. *Proc. Natl. Acad. Sci.* 104, 17152–17156. doi: 10.1073/pnas.0707678104

Thoenissen, D., Zilles, K., and Toni, I. (2002). Differential involvement of parietal and precentral regions in movement preparation and motor intention. *J. Neurosci.* 22, 9024–9034.

Tillfors, M., Carlbring, P., Furmark, T., Lewenhaupt, S., Spak, M., Eriksson, A., et al. (2008). Treating university students with social phobia and public speaking fears: internet delivered self-help with or without live group exposure sessions. *Depress. Anxiety* 25, 708–717. doi: 10.1002/da.20416

Timpe, K., and Boyd, C. A. (eds.). (2014). *Virtues and Their Vices.* Oxford: Oxford University Press.

Toni, I., Shah, N. J., Fink, G. R., Thoenissen, D., Passingham, R. E., and Zilles, K. (2002). Multiple movement representations in the human brain: an event-related fMRI study. *J. Cogn. Neurosci.* 14, 769–784. doi: 10.1162/08989290260138663

Trull, T. J. (2015). Ethical issues in researching daily life. *Monit. Psychol.* 46, 70.

Turner, J. A., Holtzman, S., and Mancl, L. (2007). Mediators, moderators, and predictors of therapeutic change in cognitive-behavioral therapy for chronic pain. *Pain* 127, 276–286. doi: 10.1016/j.pain.2006.09.005

van Oudenhoven, J. P., de Raad, B., Carmona, C., Helbig, A. K., and van der Linden, M. (2012). Are virtues shaped by national cultures or religions?. *Swiss J. Psychol.* 71, 29–34. doi: 10.1024/1421-0185/a000068

van Stralen, M. M., de Vries, H., Bolman, C., Mudde, A. N., and Lechner, L. (2010). Exploring the efficacy and moderators of two computer-tailored physical activity interventions for older adults: a randomized controlled trial. *Ann.Behav. Med.* 39, 139–150. doi: 10.1007/s12160-010-9166-8

van Stralen, M. M., de Vries, H., Mudde, A. N., Bolman, C., and Lechner, L. (2009a). The working mechanisms of an environmentally-tailored physical activity intervention for older adults: a randomized control trial. *Int. J. Behav. Nutr. Phys. Act.* 6:83. doi: 10.1186/1479-5868-6-83

van Stralen, M. M., de Vries, H., Mudde, A. N., Bolman, C., and Lechner, L. (2009b). Determinants of initiation and maintenance of physical activity among older adults: a literature review. *Health Psychol. Rev.* 3, 147–207. doi: 10.1080/17437190903229462

van Stralen, M. M., de Vries, H., Mudde, A. N., Bolman, C., and Lechner, L. (2011). The long-term efficacy of two computer-tailored physical activity interventions for older adults: main effects and mediators. *Health Psychol.* 30, 442–452. doi: 10.1037/a0023579

van Straten, A., Cuijpers, P., and Smits, N. (2008). Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. *J. Med. Inter. Res.* 10, e7. doi: 10.2196/jmir.954

Vella, J. E., Kamarck, T. W., Flory, J. D., and Manuck, S. (2012). Hostile mood and social strain during daily life: a test of the transactional model. *Ann. Behav. Med.* 44, 341–352. doi: 10.1007/s12160-012-9400-7

Villani, D., Grassi, A., Cognetta, C., Cipresso, P., Toniolo, D., and Riva, G. (2012). The effects of a mobile stress management protocol on nurses working with cancer patients: a preliminary controlled study. *Stud. Health Technol. Inform.* 173, 524–528. doi: 10.3233/978-1-61499-022-2-524

von Haaren, B., Loeffler, S. N., Haertel, S., Anastasopoulou, P., Stumpp, J., Hey, S., et al. (2013). Characteristics of the activity-affect association in inactive people: an ambulatory assessment study in daily life. *Front. Psychol.* 4:163. doi: 10.3389/fpsyg.2013.00163

Voogt, C. V., Kuntsche, E., Kleinjan, M., Poelen, E. A. P., Lemmers, L. A. C. J., and Engels, R. C. M. E. (2013). Using ecological momentary assessment in testing the effectiveness of an alcohol intervention: a two-arm parallel group randomized controlled trial. *PLoS ONE* 8:e78436. doi: 10.1371/journal.pone.0078436

Walter, K., von Haaren, B., Löffler, S., Härtel, S., Jansen, C.-P., Werner, C., et al. (2013). Acute and medium term effects of a 10-week running intervention on mood state in apprentices. *Front. Psychol.* 4:411. doi: 10.3389/fpsyg.2013.00411

Warmerdam, L., van Straten, A., Jongsma, J., Twisk, J., and Cuijpers, P. (2010). Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: exploring mechanisms of change. *J. Behav. Ther. Exp. Psychiatry* 41, 64–70. doi: 10.1016/j.jbtep.2009.10.003

Watkins, K. L., Regan, S. D., Nguyen, N., Businelle, M. S., Kendzor, D. E., Lam, C., et al. (2014). Advancing cessation research by integrating EMA and geospatial methodologies: associations between tobacco retail outlets and real-time smoking urges during a quit attempt. *Nicot. Tobacco Res.* 16, S93–S101. doi: 10.1093/ntr/ntt135

Watson, D., and Clark, L. A. (1984). Negative affectivity: the disposition to experience aversive emotional states. *Psychol. Bull.* 96, 465–490. doi: 10.1037/0033-2909.96.3.465

Watson, G. (1984). Virtues in excess. *Philos. Stud.* 46, 57–74. doi: 10.1007/BF00353491

Watts, S., Mackenzie, A., Thomas, C., Griskaitis, A., Mewton, L., Williams, A., et al. (2013). CBT for depression: a pilot RCT comparing mobile phone vs. computer. *BMC Psychiatry* 13:49. doi: 10.1186/1471-244X-13-49

Wayne, N., and Ritvo, P. (2014). Smartphone-enabled health coach intervention for people with diabetes from a modest socioeconomic strata community: single-arm longitudinal feasibility study. *J. Med. Inter. Res.* 16, e149. doi: 10.2196/jmir.3180

Webster, J. D., and Deng, X. C. (2014). Paths from trauma to intrapersonal strength: worldview, posttraumatic growth, and wisdom. *J. Loss Trauma.* 20, 253–266. doi: 10.1080/15325024.2014.932207

Weitzel, J. A., Bernhardt, J. M., Usdan, S., Mays, D., and Glanz, K. (2007). Using wireless handheld computers and tailored text messaging to reduce negative consequences of drinking alcohol. *J. Stud. Alcohol Drugs* 68, 534–537. doi: 10.15288/jsad.2007.68.534

Wenze, S. J., Armey, M. F., and Miller, I. W. (2014). Feasibility and acceptability of a mobile intervention to improve treatment adherence in bipolar disorder: a pilot study. *Behav. Modif.* 38, 497–515. doi: 10.1177/0145445513518421

Wichers, M. (2014). The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. *Psychol. Med.* 44, 1349–1360. doi: 10.1017/S0033291713001979

Wickstrom, G., and Bendix, T. (2000). The "Hawthorne effect"-what did the original Hawthorne studies actually show? *Scand. J. Work Environ. Health* 26, 363–367. doi: 10.5271/sjweh.555

Wilhelm, P., Perrez, M., and Pawlik, K. (2011). "Conducting research in daily life: a historical overview," in *Handbook of Research Methods for Studying Daily Life*, M. Mehl and T. Conner (New York, NY: Guilford), 108–123.

Williams, L. E., and Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science* 322, 606–607. doi: 10.1126/science.1162548

Wood, A. M., Froh, J. J., and Geraghty, A. W. A. (2010). Gratitude and well-being: a review and theoretical integration. *Clin. Psychol. Rev.* 30, 890–905. doi: 10.1016/j.cpr.2010.03.005

Wood, W., and Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychol. Rev.* 114, 843–863. doi: 10.1037/0033-295X.114.4.843

Wood, W., and Neal, D. T. (2009). The habitual consumer. *J. Consum. Psychol.* 19, 579–592. doi: 10.1016/j.jcps.2009.08.003

Worthington, E. L. Jr., Lavelock, C., Van, D. R., Jennings, D. J., Tongeren, A. L., Ii, G., et al. (2014). "Virtue in positive psychology," in *Virtues and Their Vices*, eds K. Timpe and C.A. Boyd (New York: Oxford University Press), 433–458.

Wrosch, C., Scheier, M. F., Miller, G. E., Schulz, R., and Carver, C. S. (2003). Adaptive self-regulation of unattainable goals: goal disengagement, goal reengagement and subjective well-being. *Pers. Soc. Psychol. Bull.* 29, 1494–1508. doi: 10.1177/0146167203256921

Yoon., K.-H., and Kim, H.-S. (2008). A short message service by cellular phone in type 2 diabetic patients for 12 months. *Diabet. Res. Clin. Prac.* 79, 256–261. doi: 10.1016/j.diabres.2007.09.007

Yüce, Y. K., Gülkesen, K. H., and Barcin, E. N. (2012). Balancing autonomy and security over geotracking patients with Alzheimer's using a personalized geotracking system with social support network. *Proc. Comput. Sci.* 10, 1064–1072. doi: 10.1016/j.procs.2012.06.150

Assessing the accuracy of self-reported self-talk

Thomas M. Brinthaupt[1]*, Scott A. Benson[1], Minsoo Kang[2] and Zaver D. Moore[1]

[1] Psychology, Middle Tennessee State University, Murfreesboro, TN, USA, [2] Health and Human Performance, Middle Tennessee State University, Murfreesboro, TN, USA

As with most kinds of inner experience, it is difficult to assess actual self-talk frequency beyond self-reports, given the often hidden and subjective nature of the phenomenon. The Self-Talk Scale (STS; Brinthaupt et al., 2009) is a self-report measure of self-talk frequency that has been shown to possess acceptable reliability and validity. However, no research using the STS has examined the accuracy of respondents' self-reports. In the present paper, we report a series of studies directly examining the measurement of self-talk frequency and functions using the STS. The studies examine ways to validate self-reported self-talk by (1) comparing STS responses from 6 weeks earlier to recent experiences that might precipitate self-talk, (2) using experience sampling methods to determine whether STS scores are related to recent reports of self-talk over a period of a week, and (3) comparing self-reported STS scores to those provided by a significant other who rated the target on the STS. Results showed that (1) overall self-talk scores, particularly self-critical and self-reinforcing self-talk, were significantly related to reports of context-specific self-talk; (2) high STS scorers reported talking to themselves significantly more often during recent events compared to low STS scorers, and, contrary to expectations, (3) friends reported less agreement than strangers in their self-other self-talk ratings. Implications of the results for the validity of the STS and for measuring self-talk are presented.

Keywords: Self-Talk Scale, self-reports, inner experience, personality assessment, individual differences

Introduction

Conducting research on the psychology of inner experiences is an interesting and challenging activity. Because the phenomena of interest may be covert, hidden, or completely unobservable by an outside agent, researchers must rely primarily on the introspection and self-reports of participants. Several of the other papers in this special issue address ways to overcome some of the limitations of and provide complements to self-report with respect to different kinds of inner experiences. In the present paper, we describe three studies designed to assess the accuracy of self-reported self-talk.

There is a long history of research interest in the veridicality of self-reports (e.g., Schoeneman, 1981; Shrauger and Osberg, 1981; Moskowitz, 1986; Brinthaupt and Erwin, 1992; Vazire and Wilson, 2012). Interest in the accuracy of self-reports covers a broad range of phenomena, particularly behaviors that might be expected to show socially desirable responding effects. For example, Gatersleben et al. (2002) found that self-reports of pro-environmental behaviors were only weakly related to actual household energy use. In a meta-analysis of the validity of self-reported drug use, Magura and Kang (1996) found that underreporting was a prevalent issue. People have also been found to underestimate their frequency of sedentary behaviors

(e.g., Klesges et al., 1990) and underestimate their dietary intake (e.g., Palaniappan et al., 2003).

With respect to reporting about specific self-related phenomena, the literature finds that self-reports of personality traits generally show moderate agreement with observations from others and behavioral indicators (e.g., Mehl et al., 2006; Vazire and Mehl, 2008; Back et al., 2009). Research also suggests that people are more accurate when reporting about personality disorders than when reporting about mental disorders, possibly because the former are seen as part of one's self-definition and as not as something that is unacceptable or a reflection of a disturbance (e.g., Oltmanns and Turkheimer, 2006). Furthermore, in domains that reflect more positive personal behaviors or characteristics, research suggests that when people are uncertain, they are likely to report or demonstrate having personality traits that are seen by others as positive, such as prosocial characteristics (Chin et al., 2012).

Research on maladaptive cognitive and affective variables supports the idea that self-reports are generally similar to the reports of knowledgeable informants (e.g., South et al., 2011), although some clinical issues, such as narcissistic personality disorder, appear to be particularly susceptible to self-other disagreement (Klonsky et al., 2002). Carlson et al. (2013) found that self-reports showed greater validity than informant reports for internalizing personality disorders (such as showing neuroticism, anxiety, or obsessive–compulsive tendencies). Alternatively, they found that informant reports showed greater validity than self-reports for personality disorders that are more externalizing in nature (e.g., being disagreeable, aggressive, or narcissistic).

Whereas self-talk appears to be primarily an internalizing rather than externalizing phenomenon, it is unclear how people perceive the appropriateness of their self-talk. Perceptions of self-talk appropriateness are likely to differ depending on whether we consider the frequency of self-talk (e.g., how often people talk to themselves) or its affective (i.e., positive and negative) content. Researchers find that people are frequently biased toward self-enhancing self-perceptions, especially individuals who show high levels of narcissism (Schriber and Robins, 2012). However, it is unclear the extent to which self-talk is seen as a socially desirable or undesirable characteristic. Brinthaupt et al. (2009) found that self-reported self-talk was only weakly related to a measure of social desirability. That finding suggests that self-talk frequency may not be seen in a particularly negative way among respondents.

Maladaptive or dysfunctional self-talk content – unrealistic, irrational, or excessively negative – has been a focus of cognitive-behavioral therapists for many years (e.g., Beck, 1976; Glass and Arnkoff, 1994). An implication of this focus is that certain kinds of self-talk may be seen by people as less socially desirable. If this is the case, then the accuracy of certain kinds of self-reported self-talk, such as self-critical self-statements, may be negatively associated with perceptions of inappropriateness or social undesirability. More positive or affectively neutral self-talk, such as self-managing self-statements, should be less strongly related to those perceptions. People's reports of the frequency of different kinds of self-talk might therefore be affected by their beliefs or presuppositions about how maladaptive or dysfunctional it is (Hurlburt and Heavey, 2015).

Self-talk frequency is related to a wide variety of self-regulatory behaviors (e.g., Mischel et al., 1996; Carver and Scheier, 1998; Leary, 2004). With the increasing interest in self-talk as a psychological phenomenon across multiple domains (Beck, 1976; Kendall et al., 1989; Hardy et al., 2009; Winsler et al., 2009; Hurlburt et al., 2013), it is crucial that data be collected that examine the accuracy of self-reported self-talk. Among the important questions here include (1) the extent that people's reports of their self-talk frequency correspond to actual behavioral instances of self-talk across a variety of everyday situations or circumstances and (2) whether people's awareness of their self-talk reflects their self-reported frequency of self-talk instances, as assessed at different times, or through different kinds of data collection.

Self-Talk and the Self-Talk Scale

The development and initial validation of the Self-Talk Scale (STS; Brinthaupt et al., 2009) permits researchers to examine individual differences in self-talk frequency. The STS is general measure of self-talk that is applicable to a broad range of self-regulatory behaviors and situations. The scale consists of 16 items rated on a five-point frequency scale (1 = *never*, 5 = *very often*) using the common stem "I talk to myself when. . ." Brinthaupt et al. (2009) showed that the STS has a structure consisting of one higher-order factor (overall self-talk) and four primary factors (self-critical, self-reinforcing, self-managing, and social-assessing self-talk), along with acceptable test–retest stability and internal consistency.

Self-critical self-talk is generally associated with negative events (e.g., "I feel ashamed of something I've done" or "Something bad has happened to me"). Self-reinforcing self-talk focuses on positive events (e.g., "I'm really happy for myself" or "I'm proud of something I've done"). Self-managing self-talk pertains to general self-regulation (e.g., "I'm mentally exploring a possible course of action" or "I'm giving myself instructions or directions about what I should do or say"). Finally, social-assessing self-talk refers to people's social interactions (e.g., "I want to replay something that I've said to another person" or "I want to analyze something that someone recently said to me").

In their STS validation work, Brinthaupt et al. (2009) found negative relationships between social-assessing and self-critical self-talk and self-esteem, as well as a positive association between self-reinforcing self-talk and self-esteem. Frequent self-talkers (i.e., those scoring in the upper quartile on the total STS) also scored higher than infrequent self-talkers (lowest quartile) on need for cognition and obsessive–compulsive tendencies.

Additional research using the STS shows more frequent self-talk among adults who reported having had an imaginary companion in childhood or who grew up as an only child compared to those who did have those experiences (Brinthaupt and Dove, 2012) and a high negative correlation between loneliness and mental health in more frequent self-talkers (Reichl et al., 2013). Research using an adaptation of the STS (Shi et al., 2015) has found that individuals with high public speaking anxiety were

cognitively "busier" (i.e., reported higher levels of several kinds of self-talk) than those with low anxiety as they prepared for an upcoming speech. Finally, research has supported the use of the STS response format and the use of the STS total score as a unidimensional measure of self-talk frequency (Brinthaupt and Kang, 2014), as well as shown a similar factor structure in a cross-cultural comparison (Khodayarifard et al., 2014).

In summary, the STS possesses good psychometric properties, and scores on the measure have been associated with a wide range of interesting phenomena. Clearly, the scale shows promise as a measure of self-reported self-talk. However, an important question is to what extent the self-talk that STS respondents report reflects their actual self-talk frequency. As Hurlburt and Heavey (2015) have shown, respondents' reports of any kind of inner experience can be problematic, particularly when those reports are retrospective. Previous reviews of self-talk measures (e.g., Glass and Arnkoff, 1994; Uttl et al., 2011) propose that validity can be examined through the use of multiple measures and assessment occasions. There are several interesting theoretical and research implications related to the analysis of the accuracy of self-reported self-talk and to the different self-talk functions. In this paper, we report the results of three studies examining the correspondence between self-reports of self-talk using the STS and behavioral or observer indicators of self-talk. In addition, we examine the four STS sub-scales and how these are associated with different levels of accuracy or agreement.

Study 1: STS Scores and Recent Self-Talk Experiences

As with other domains of personality self-knowledge (Back and Vazire, 2012), the accuracy of self-reports of one's self-talk frequency is likely to be affected by explicit and implicit information processing, the salience of relevant behavioral instances, accessibility to ongoing inner experiences, and information from other people (Hurlburt et al., 2013). It is likely that people rely on a wide variety of information sources and presuppositions when completing the STS. We would expect that respondents who notice or recall more situations in which they have talked to themselves in the past should be likely to report more frequent self-talk. Assuming that people respond to the STS based on an aggregation of behavioral instances, cognitive heuristics, presuppositions, times, and situations associated with self-talk (Kenrick and Funder, 1988; Hurlburt et al., 2013), self-talk likelihood in sample situational instances should be positively associated with self-reports of typical self-talk.

In Study 1, participants first completed the STS and then, at a later time, completed a revised version of the measure. We expected that, if the STS assesses respondents' awareness of or assumptions about their typical self-talk frequency, then high STS scorers would report being more likely than low STS scorers to talk to themselves when specific STS-related situations occur. Thus, we predicted that previous overall and subscale STS scores, based on how often people report typically talking to themselves, would be positively and significantly correlated with their reports of the frequency of self-talk in response to relevant situations that had recently occurred.

Method
Participants

Through the cooperation of three faculty members from human sciences, speech, and English departments, 83 students (27 men, 56 women) from a large southeastern U. S. public university completed the materials. Students' ages ranged between 18 and 31 years ($M = 20.01$, SD $= 2.23$). For two of the instructors, students completed the surveys during their normal class time, with the other instructor permitting students to complete the measures outside of class and returning them at the next class period. Students volunteered to participate, completed an informed consent form, and received a small number of extra credit points for their participation at the discretion of their instructors. This study (as well as Studies 2 and 3 reported later) received IRB approval prior to data collection.

Materials

Students completed the original STS early in the academic term and then, approximately 6 weeks later, completed a revised version of the measure. STS internal consistency coefficients were acceptable for the overall scale ($r = 0.92$) and the subscale (rs ranging from 0.79 to 0.87). For the modified, "recent experience" version of the STS (reSTS), participants indicated whether they had recently experienced any of the 16 situations from the original STS, as well as whether they had engaged in self-talk during those experiences. The reSTS wording of the items was modified from the original STS to reflect the past-tense nature of the question posed. For example, we changed the original STS item "I want to replay something I've said to another person" to "I wanted to replay something I said to another person" for the reSTS. Wording was only changed as much as necessary to make the items grammatically correct in the past tense. The order of the items in the reSTS was the same as the original STS.

Participants received the following instructions for completing the reSTS: "check all that apply; over the past 2 days, I have been in a situation where. . ." They rated the 16 items corresponding to the STS items (e.g., "Something good happened to me" and "I was really upset with myself") in terms of whether that situation had occurred (yes/no). For each situation that had occurred, students then indicated (yes/no) whether they had talked to themselves "(either silently or aloud) about that situation as it occurred or shortly after it occurred." From these data, we calculated the total number of STS-related situations that had occurred over the past 2 days (possible range 0–16), the number of situations that had occurred for each of the four STS subscales (possible range 0–4), and the total and subscale ratios of situations where students reported talking to themselves when those situations had occurred (possible range 0.000–1.000). For the latter subscale ratios, when no instances of subscale situations occurred, values were set to missing.

Procedure

As noted earlier, participants received materials twice in a semester during regular class time. At the first session, they

received a consent form, a paper copy of the STS, and a demographic form that included questions about age and gender. Students also included their student ID numbers on the form, in order to permit matching of data from the second session. At the second data collection period, we visited the same classrooms, and the participating students completed the reSTS on paper. After the data from the second session were collected, we thanked and debriefed the participants. The surveys, including the consent form, took no more than 10 min to complete.

Results and Discussion

Table 1 provides the descriptive statistics for the reSTS measures. These data indicate that, in the 2-day period used for assessment, approximately 67% of the 16 STS situations had occurred. Of these situations, self-managing ones were most frequent, whereas self-critical were least frequent. The data also indicated that participants reported self-talk associated with over 80% of STS situations that had occurred. The self-talk ratio was highest for the self-managing and lowest for the self-reinforcing situations. These data support the idea that there are numerous daily opportunities for people to talk to themselves and that they frequently report doing so when they encounter those situations.

Table 2 provides the correlations among the major measures for both testing sessions. As the table indicates, STS total scores were positively correlated with the reSTS overall ratio. In other words, frequent self-talkers reported talking to themselves more often than infrequent self-talkers when STS-related situations recently occurred. Each of the STS subscales were also positively and significantly related to their corresponding reSTS ratio scores, with the self-critical and self-reinforcing scores showing stronger relationships than the self-managing and social-assessing scores. In addition, STS total scores were positively related to the total number of STS situations that had occurred in the previous 2 days, $r(81) = 0.395$, $p < 0.001$.

These data provide good support for the expectation that typical self-reported self-talk scores, assessed 6 weeks earlier, would correspond to self-reports of recent examples of instances where self-talk could occur. There are at least two possible explanations for these results. First, general assessments of people's typical self-talk frequency are an accurate representation of the frequency with which they actually talk to themselves across a variety of specific and recent self-regulatory situations. For example, people who report that they frequently talk to themselves when something bad or good happens to them also report having done so when something bad or good has recently happened. Second, it is possible that people customarily recall salient examples of self-talk (e.g., within the previous few days) when they assess the frequency of their typical self-talk using the STS. If this recall pattern occurs, then we have simply assessed the situations that respondents are already relying on when they rate their typical self-talk patterns using the original STS.

The fact that STS scores were significantly associated with the occurrence of STS-related situations suggests that frequent self-talkers are more aware of or responsive to a variety of self-regulatory situations than infrequent self-talkers. Perhaps self-talk frequency is positively associated with a greater overall sensitivity to one's intrapersonal and interpersonal experiences. Of course, we cannot determine the extent to which participants were recalling their previous STS responses when rating their self-talk in recent situations. If they were able to recall their earlier STS responses and were motivated to be consistent, then this could account for the observed relationships. Future research could address these possible explanations.

In summary, Study 1 provided data to suggest that (1) the self-talk situations included in the STS are frequently reported occurrences in people's lives, (2) self-talk is reported to frequently occur in those situations, and (3) respondents' self-reports of their typical self-talk generally agree with their reports about specific related experiences. Although these results were encouraging, a variety of alternative explanations could not be addressed or controlled in this study.

Study 2: Experience Sampling Study of Current Ongoing Self-Talk

A limitation of Study 1 is that participants' recall of relevant instances and their accompanying self-talk was based on events from the past 48 h. Because of the time lag involved, the extent

TABLE 1 | Study 1: Descriptive statistics for reSelf-Talk Scale (STS) measures.

Item	Mean	SD	*n*
reSTS event frequency			
Overall	10.78	3.66	83
Self-critical	2.04	1.44	83
Self-reinforcing	2.70	1.36	83
Self-managing	3.10	1.10	83
Social-assessing	2.95	1.20	83
reSTS self-talk ratio			
Overall	0.815	0.249	83
Self-critical	0.782	0.363	71
Self-reinforcing	0.695	0.394	74
Self-managing	0.870	0.272	83
Social-assessing	0.810	0.334	80

reSTS refers to revised STS items; event frequency refers to the number of STS-related situations (out of 16 for overall, out of four for each subscale) that had occurred; self-talk ratio refers to the proportion of instances of STS situations that had occurred in which participants reported that associated self-talk had also occurred.

TABLE 2 | Study 1: Correlations among the STS and reSTS measures.

STS:	Total	Self-critical	Self-reinforcing	Self-managing	Social-assessing
reSTS ratio					
Overall	0.446***	0.331**	0.356***	0.460***	0.372***
Self-critical	0.377***	0.373***	0.231	0.332**	0.372***
Self-reinforcing	0.412***	0.202	0.439***	0.417***	0.340**
Self-managing	0.231*	0.134	0.170	0.257*	0.225*
Social-assessing	0.323**	0.221*	0.218	0.371***	0.293**

*p < 0.05, **p < 0.01, ***p < 0.001; reSTS ratio refers to the proportion of instances of STS situations that had occurred in which participants reported that associated self-talk had also occurred.

of agreement between original STS scores and recent instances of self-reported self-talk may have been constrained. A stronger test of the self-report accuracy question would involve very recent experiences that should be more salient and accessible to the participants.

The experience sampling method (ESM) has proven to be a reliable and valid method for measuring a wide range of inner experiences (Csikszentmihalyi and Larson, 2014). In Study 2, we utilized ESM methods to compare general STS reports with the self-talk that is reported to occur in response to current experiences. In particular, we examined the self-talk patterns of those who fell in the upper and lower quartiles of the STS total scores. These self-reported *frequent* and *infrequent* self-talkers were prompted periodically on their smart phones for 5 days to indicate whether any of the 16 STS situations had occurred within the past 2 h. If so, they then reported whether or not they had talked to themselves about that instance.

We expected that, when instances of STS-related situations occurred, previously identified frequent self-talkers would report more accompanying self-talk than would the infrequent self-talkers. In other words, across the array of situations included in the STS, more overall and subscale self-talk should be reported during those situations for the frequent than the infrequent self-talkers.

Method
Participants
Using data collected approximately 1 month earlier from the university's General Psychology pretesting research pool, we recruited 35 participants (8 male, 27 female) from the upper ($n = 20$) and lower ($n = 15$) quartiles of total STS scores. The lower quartile participants ($M = 19.93$, SD $= 6.46$) differed significantly on STS scores from the upper quartile participants ($M = 50.70$, SD $= 5.39$), $t(33) = 15.34$, $p < 0.001$. Participants ranged in age from 18 to 26 years ($M = 18.66$, SD $= 1.78$). With respect to ethnicity, 74% of the participants were Caucasian and 11% were African–American. In order to participate in the study, students were required to possess an operable smart phone that could receive text messages and connect to the Internet.

Materials and Procedure
Using a modified version of the STS, we asked participants about their very recent or currently ongoing activities and whether those activities were associated with a self-reported instance of self-talk. Due to software limitations, the order of the questions was the same as the original STS. We modified the STS items in two key ways for this study. First, each item included the phrase "Over the last *two hours*, I have been in a situation where…" Care was taken to change only what was necessary to match tense for the complete phrase. Second, participants simply answered *yes* or *no* to whether each of the 16 situations had recently occurred. For items answered *yes*, a follow-up yes/no question appeared: "Did you talk to yourself (either silently or aloud) during or immediately after the situation occurred?"

We used a commercial survey hosting web-service to administer materials and organize response data. A free Gmail account

and a paid account to Right Inbox scheduled the text messages. Right Inbox is a web-browser extension that allows e-mail drafts in the Gmail web-based e-mail client to automatically send themselves at scheduled times. Over a 5-day period, participants received 25 text message prompts (five each day) during a 10-h daily period (10 am–8 pm). Participants used the same assigned number to identify themselves at the start of each of the surveys.

There were 32 possible yes/no questions on each survey text prompt. Possible scores on these measures ranged between 0–16 for the first questions (situations that had recently occurred) and the follow-up (talking to oneself if the situation had occurred). Thus, over the course of the study, there were 400 possible instances where an STS situation could have occurred and where participants could have reported talking to themselves. As in Study 1, we calculated the total and subscale ratios of situations where students reported talking to themselves when those situations had occurred (possible range 0.000–1.000). For the latter subscale ratios, when no instances of situations occurred, values were set to missing.

A random number generator provided scheduled times for contacting research participants. Twenty-five numbers between 0 and 119 were selected at random, and the corresponding number of minutes was added to the start time for each of the five 2-h blocks. For example, a 63 selected for a 10 am–12 pm block would add 63 min to the 10 am start time, so that the text message would be scheduled to be sent at 11:03 am. In the selection of these contact times, we ensured that no two consecutive prompt times occurred within 30 min of each other.

E-mails were converted to text messages using standardized e-mail addresses issued by cell phone carriers to each phone number. Standard format for an assigned "e-mail to SMS" e-mail address is the 10 digits of the phone number "at" a domain hosted by the carrier. For example, a Verizon phone number, (931) 555-1234, can receive as text messages any e-mail sent to 9315551234@vtext.com. We recorded the appropriate text message addresses for all research participants and sent a pilot message before the study began in order to identify and correct any address issues.

Each unique link that led to the survey directed participants to a new "collector," or a new identifiable instance of the survey. This allowed data from each of the 25 surveys to remain separate from one another, while the participant's ID number allowed us to collate these data. Additionally, time signatures on each collector allowed for sorting the data by time-order.

At an orientation session at the start of the study, we briefed participants on the details of the ESM, including the scheduling of the text messages and that surveys would need to be completed online and within 2 h of receiving the link to each survey. Each participant provided their phone number and carrier for receiving text messages. After initial testing of the text message system and phone compatibility, students received links to surveys through text messages, beginning the following Monday. Participants agreed to receive five text messages a day for five consecutive days. All students participated during the same 5-day period and received all text messages at the same intervals.

The links in the text messages directed participants to a survey hosted on a commercial survey website. Surveys remained open

for 2 h after receiving the text message link. This gave participants some flexibility in answering the survey without allowing excessive overlap of reporting periods. After 2 h, the link instead directed participants to a page explaining that the survey was closed.

Over the 2 weeks following the testing week, participants returned for debriefing and to receive credit for participating in the study. The exit survey included demographic items (ID number, age, gender, and ethnicity) and a yes/no question about whether they had ever considered their level of self-talk prior to the study. In addition, participants rated seven items regarding their experiences during the ESM study, using a five-point scale (1 = very little, 5 = very much). These items included the difficulty in determining and recalling whether self-talk had occurred when prompted, the difficulty of completing the survey on time, whether the study questions and directions were clear, whether the survey responses captured most daily instances of self-talk, and whether their awareness of self-talk increased after participating in the study.

Results and Discussion

Results showed that participants took an average of 20.32 min (SD = 24.74) to respond to the receipt of the text messages. Frequent and infrequent self-talkers did not differ significantly on this measure [$t(34) = 0.186$, $p = 0.85$]. Participants reported that approximately 23% of the 400 possible STS-related situations had occurred (M = 91.46, SD = 47.02) over the 5-day period. In addition, participants reported talking to themselves 65% of the time when the overall STS-related situations had occurred. With respect to the STS subscales, participants reported talking to themselves 72% of the time for the self-critical and self-managing situations, 63% for the social-assessing situations, and 51% for the self-reinforcing events.

Data for the major measures by frequent and infrequent self-talkers are presented in **Table 3**. As the table shows, the groups did not differ significantly on the number of STS-related situations that had occurred over the 5 days or the number of

TABLE 3 | Study 2: Comparison of infrequent and frequent self-talkers on major measures.

	Infrequent (n = 15)		Frequent (n = 20)			
	Mean	**SD**	**Mean**	**SD**	**t**	**p**
# of STS situations	81.73	54.92	98.75	40.05	1.061	0.296
# of situations w/self-talk	53.60	54.37	76.85	43.14	1.411	0.167
Self-Talk Ratio						
Overall	0.538	0.253	0.734	0.216	2.465	0.019
Self-critical	0.632	0.269	0.779	0.236	1.722	0.094
Self-reinforcing	0.355	0.320	0.619	0.221	2.889	0.007
Self-managing	0.598	0.317	0.808	0.193	2.431	0.021
Social-assessing	0.586	0.371	0.655	0.339	0.558	0.581

Self-talk ratio refers to the proportion of instances of STS situations that had occurred in which participants reported that associated self-talk had also occurred. Reported ratio values reflect the average of the proportions for the individual members of each group.

those situations in which self-talk was reported to have occurred. These results suggest that the everyday experiences related to the topics included in the STS are similar for frequent and infrequent self-talkers. However, as expected, frequent self-talkers reported a significantly higher overall proportion of talking to themselves when STS situations had occurred than did infrequent self-talkers. Examination of the STS subscales showed that frequent self-talkers differed significantly from infrequent self-talkers in their reports of self-reinforcing and self-managing self-talk. Infrequent self-talkers reported being least likely to talk to themselves during self-reinforcing situations, whereas frequent self-talkers reported being most likely to talk to themselves in response to self-managing situations.

The final analyses examined the post-study survey data and how participating in the study affected people's attention to their self-talk. Twenty-seven of the 35 participants (13 infrequent, 14 frequent self-talkers) completed this survey. With respect to having ever considered their level of self-talk prior to the study, 67% of the infrequent self-talkers reported no and 62% of the frequent self-talkers reported yes [$X^2(1) = 1.99$, $p = 0.158$]. The two groups did not differ significantly on any of the survey measures. Thus, the participants' reports of their study experiences were similar regardless of their self-talk frequency status.

In summary, there was little evidence that the frequent and infrequent self-talkers differed in how often STS-related situations occurred during the 5 days of the study or that the two groups experienced the study methodology differently. However, frequent self-talkers reported being significantly more likely to talk to themselves when those situations occurred than did the infrequent self-talkers. This result provides additional support for the validity of the STS and of individuals' self-reports of their typical self-talk frequency.

It is interesting that the infrequent self-talkers (as identified by their scores on the STS from the previous month) reported talking to themselves when STS situations occurred nearly 54% of the time. This is higher than the percentage of self-talk frequency based on this group's mean STS scores (19.93/64 or 31%). The frequent self-talkers reported talking to themselves when STS situations occurred 73% of the time, which is very similar to their mean STS score percentage (50.70/64 or 79%). It appears that those who rate themselves as infrequent self-talkers still report talking to themselves around half of the times shortly after STS-related situations occur, but that they under-report that frequency when completing the STS. It is also possible that there was a regression to the mean effect for the infrequent and frequent self-talkers. Because we selected lower- and upper-quartile STS scorers for the groups, their later situation-specific self-talk reports might be more likely to increase (decrease).

Our methodology in Studies 1 and 2 did not permit an assessment of the amount of time spent talking to oneself when STS situations occurred. We only asked participants to indicate whether they had talked to themselves in response to the situation occurring. Future research could examine the length, depth, or salience of one's self-talk following the occurrence of these situations. Frequent self-talkers would be expected to show large

differences on these self-talk characteristics compared to infrequent self-talkers. It is possible that degree of cognitive processing of events contributes more to people's general assessments of their self-talk frequency (and whether they are categorized as frequent or infrequent self-talkers) than the occurrence of situations that prompt self-talk.

The first two studies provided good support for the prediction that self-reports of typical self-talk frequency using the STS are accurate. Analysis of the STS subscales revealed some interesting qualifications to the general trend. Study 1 showed that all the STS subscales were significantly correlated with recent situations (which had occurred within the past 2 days). However, Study 2 showed that frequent and infrequent self-talkers did not differ in their self-reported self-critical and social-assessing self-talk frequency in specific situations that had occurred very recently. It is possible that these categories of self-talk are more difficult to estimate accurately. It is noteworthy that both of these subscales assess more negative than positive self-talk instances (Brinthaupt et al., 2009). Individuals appear to be most accurate in judging their typical self-reinforcing self-talk frequency. Once again, research examining the length, depth, or salience of one's self-talk would provide important information about the memorability of different kinds of self-talk and how that memorability contributes to assessments of one's typical self-talk frequency.

Study 3: Self and Other Ratings of Self-Talk Frequency

Another way to examine the validity of self-reported self-talk is to compare self-reports to the reports of knowledgeable others. Unfortunately, there are several reasons why self-talk might not be easy to monitor in another person. This task may be similar to comparing self- and other-reports of a person's internal physiological states (e.g., the severity of a headache and one's hunger status). In these cases, there is very little information that an observer could rely on to provide an accurate assessment. In addition, due to self-presentation or impression management reasons (such as concerns about one's "sanity"), both silent and aloud self-talk are probably more likely to be used when a person is alone than with others. By its very nature, self-talk is self-directed speech that appears to not be intended for the ears of other people. Due to issues of attention or focus, it may also be more difficult mechanically and socially to engage consistently in self-talk in the presence of another person than when alone. Thus, the ability for an observer to assess accurately the self-talk frequency of a target person may be fundamentally limited.

Despite the intrapersonal nature of self-talk, there is evidence that there are interpersonal aspects of the phenomenon. For example, research shows that children (and to a lesser extent adults) will engage in more private speech when performing tasks in the presence of others than when alone (McGonigle-Chalmers et al., 2014). There may be times when people talk to themselves (either silently or aloud) in the presence of others for strategic reasons (e.g., to convey an emotional response or to indicate that

one is actively thinking about an issue). Thus, there appear to be some interpersonal aspects of the inner experience of self-talk. If this is true, then other people, particularly those with extensive knowledge of and experience with the respondent, should show high levels of agreement with that respondent's self-reported self-talk frequency.

There is research support for the idea that other people can contribute to the accuracy of one's self-views. For example, Srivastava (2012) noted that personal attributes that are reputational in nature (such as social status or likeability) are strongly affected by how one is seen by others. Because covert or silent self-talk is a highly internalized phenomenon, these results suggest that self-reports of self-talk might be more valid than the reports of informants. On the other hand, people who engage in frequent private speech (out loud self-talk) may cue informants about their likely inner speech (silent self-talk) frequency.

In the realm of personality pathology, South et al. (2011) found that, with a community sample of married couples, as degree of acquaintance increased, self-other agreement about extent of pathology increased. As noted earlier, there is also a research literature on the tendency for individuals to rate themselves higher on maladaptive cognitive and affective variables than do those who know them well. For example, there is a bias toward more favorable (less negative) other-reports than self-reports in the domain of personality pathology (Vazire, 2010).

If self-talk is seen as a maladaptive behavioral characteristic, then we would expect that people will self-report higher levels of self-talk than what they will report for their partners. However, as noted earlier, Brinthaupt et al. (2009) found that scores on the STS were only weakly related to social desirability scores. In addition, research shows that people generally use less private speech and more inner speech as they move from childhood to adulthood (Winsler and Naglieri, 2003; Duncan and Tarulli, 2009). Thus, adult self-talk tends to be hidden from others, which should also create a tendency toward higher self-reported than other-reported self-talk scores.

In Study 3, we examined the extent to which the reports of others correspond with self-reported self-talk frequency. For this study, pairs of participants, who either knew their partner well or did not know their partner at all, rated themselves and the other on the STS as well as a measure of private speech (out loud self-talk). The stranger data served as the control condition, providing a baseline for what respondents think about how often people in general normally talk to themselves. We expected that close others would show greater self-other agreement on self-talk frequency than would strangers. We also expected that, for those who know each other, increased relationship closeness would be associated with increased levels of self-talk frequency agreement. These predictions are based on the assumption that greater relationship closeness will provide partners with more situations where self-talk might occur and more accurate information about how often self-talk occurs, compared to strangers. Finally, because of its greater observability, we expected stronger self-other agreement results for private speech (i.e., out loud self-talk) than for self-talk measured with the STS (i.e., both inner and private speech).

Method

Participants

Eighty-eight students (44 pairs) participated in the study. Participants were drawn from the department's General Psychology research pool. They received course credit for their participation. The sample includes 29 men and 59 women, with an average age of 19.75 years (SD = 1.80). With respect to ethnicity, 56% were Caucasian and 35% were African–American. There were 26 pairs in the friends group (18 men, 34 women) and 18 pairs (11 men, 25 women) in the stranger group. Eighteen (69%) of the friend pairs were same-sex and nine (50%) of the stranger pairs were same sex. Forty-two (81%) of the friend group participants identified themselves as friends, siblings, or roommates, with nine participants (17%) indicating an exclusive or non-exclusive dating relationship.

Measures

Participants completed the STS and a measure of out loud private speech for both themselves and another person, as well as demographic items and partner ratings. The private speech measure was Duncan and Cheyne's (1999) Self-Verbalization Questionnaire (SVQ). This is a 27-item measure of activities and situations in which people might talk out loud to themselves. The SVQ consists of four factors, including spatial-search (e.g., "I sometimes verbalize my thoughts when I'm searching for a book in a library"), behavioral-organizational (e.g., "I sometimes think out loud to myself when I'm trying to clean up a mess in a big hurry"), cognitive-attentional (e.g., "I sometimes verbalize my thoughts when I'm memorizing something for an exam"), and affective (e.g., "I sometimes verbalize my thoughts when I'm feeling angry or upset about something"). Respondents rate the items using a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*). Items are summed to create the subscale and total scores, with possible total scores ranging from 27 to 189. We used only the total scores for this study. The authors report acceptable reliability and validity for the SVQ. In the current sample, the alpha coefficients for the overall SVQ were in the acceptable range for both self-ratings (0.92) and other-ratings (0.93). The alpha coefficients for the overall STS were also in the acceptable range for both self-ratings (0.89) and other-ratings (0.88).

To determine the accuracy of respondents' self-reports, we created an absolute percentage error (ape) score for the STS and SVQ. The apeSTS was calculated as the absolute value of the difference between the partner's rating of the participant's STS score and the participant's self-reported STS score, divided by the participant rating, and then multiplied by 100. The apeSVQ score was calculated as the absolute value of the difference between the partner's rating of the participant's SVQ score and the participant's self-reported SVQ score, divided by the participant rating, and then multiplied by 100. A smaller ape score represented greater similarity in the ratings of the two partners. Similar measures have been used by researchers to assess measurement accuracy in other domains (e.g., Kang et al., 2012).

For the two self-talk measures, we also calculated a measure of bias – the difference between the partner's rating of the participant's STS or SVQ score and the participant's self-reported STS or SVQ score, divided by the participant rating, and multiplied by 100 (biasSTS and biasSVQ, respectively). Bias scores in the positive direction indicated a bias toward reporting higher self-talk scores for the other than for self. Finally, we calculated a simple difference score for each measure – the difference between the partner's score for the participant and the participant's STS or SVQ score (diffSTS and diffSVQ, respectively). Difference scores in the positive direction reflected higher other-reported than self-reported self-talk scores.

Demographic items included participant age, gender, and ethnicity. Participants in the significant other condition indicated the nature of the relationship with their friend/partner (e.g., friend, roommate, dating), whether they currently lived with this person, and how long they had known them. Next, they indicated how much time in an average day and an average week they spent in the physical presence of their partner. Finally, participants in both the significant other and stranger conditions rated, using five-point scales, how close they and their partner were (1 = *not close at all*, 5 = *extremely close*), how well they knew their partner (1 = *not very well at all*, 5 = *extremely well*), how well they understood how their partner thinks about him/herself (1 = *not very well at all*, 5 = *extremely well*), how well they would say that their partner knows them (1 = *does not know me very well at all*, 5 = *knows me extremely well*), and how well they would say that their partner understands how they think about themselves (1 = *not very well at all*, 5 = *extremely well*).

Procedure

Prior to the testing session, we informed participants in the significant other condition to bring a friend, romantic partner, or close other with them in order to participate and receive course credit. Participants in the stranger condition were randomly assigned a partner at the start of the testing session. We collected data for the two conditions in separate testing sessions, with all participants in each session falling in the same condition. Participants first completed the self-talk/private speech measures for themselves and their partner (in random order), followed by the demographic and other items.

Results and Discussion

Descriptive Statistics

Participants in the friend group differed significantly from participants in the stranger group in the expected direction on all of the relationship measures. For example, compared to the stranger participants, the friend participants reported being closer to their partners [M_F = 4.12, SD_F = 0.96; M_S = 1.08, SD_S = 0.28; $t(86)$ = 18.33, p = 0.000] and knowing their partners better [M_F = 3.98, SD_F = 1.02; M_S = 1.00, SD_S = 0.00; $t(86)$ = 17.52, p = 0.000], as well as their partners knowing them better [M_F = 3.92, SD_F = 0.98; M_S = 1.03, SD_S = 0.17; $t(86)$ = 17.32, p = 0.000].

We examined the correlations among the STS and SVQ ratings for the entire sample. Self-rated STS scores were highly correlated with participants' ratings of their partners, $r(86)$ = 0.688, p < 0.001. Similarly, self-rated SVQ scores were highly correlated with participants' ratings of their partners, $r(86)$ = 0.695, p < 0.001. These results suggest that participants assumed that their own levels of self-talk were similar to the levels likely to

be shown by their partners. In addition, self-rated STS and SVQ scores were significantly correlated, $r(86) = 0.503$, $p < 0.001$, as were other-rated STS and SVQ scores, $r(86) = 0.608$, $p < 0.001$. These results indicate substantial overlap between the two self-talk measures.

The apeSTS scores for the entire sample ($M = 21.63$, $SD = 21.74$) differed significantly from 0 (perfect accuracy), $t(87) = 9.33$, $p < 0.001$, as did the apeSVQ scores ($M = 23.86$, $SD = 19.95$), $t(87) = 11.22$, $p < 0.001$. These results indicated that participants tended to rate themselves differently on self-talk compared to how their partners rated them. The bias and difference scores for the entire sample did not differ significantly from 0.

Comparison of Friend and Stranger Conditions

Table 4 presents data for the self-talk measures. As the table indicates, comparison of the two groups on the accuracy measures revealed one significant difference – the friend group showed larger average percentage error scores than the stranger group for the STS. This finding was opposite to what we expected. Because we failed to find any differences on the bias or difference measures, the apeSTS results suggest that participants in the friend condition were generally less accurate in rating their partner's self-reported self-talk frequency than those in the stranger condition, but not in a specific (over- or under-reporting) direction. Separate paired-sample t-tests for each group indicated that the friend group did not differ in their self- and other-ratings on the STS or SVQ. However, the stranger group reported significantly higher self-ratings ($M_s = 58.19$, $SD_s = 8.44$) than other-ratings ($M_o = 54.81$, $SD_o = 7.36$) for the STS, $t(35) = 2.66$, $p = 0.012$. Strangers also reported significantly higher self-ratings ($M_s = 125.44$, $SD_s = 21.74$) than other ratings ($M_o = 117.42$, $SD_o = 21.22$) for the SVQ, $t(35) = 2.66$, $p = 0.012$. Thus, there was a tendency toward more frequent self-rated self-talk than other-rated self-talk for the strangers but not for the friends.

TABLE 4 | Study 3: Comparison of friend and stranger groups on major measures.

	Friend (n = 52)		Stranger (n = 36)			
	Mean	SD	Mean	SD	t	p
STS Self	57.54	12.52	58.19	8.44	0.274	0.785
STS Other	56.94	11.87	54.81	7.36	0.959	0.340
SVQ Self	123.19	29.56	125.44	21.74	0.390	0.698
SVQ Other	122.88	28.73	117.42	21.22	0.972	0.334
apeSTS	25.76	26.19	15.65	10.65	2.192	0.031
biasSTS	4.79	36.60	−3.80	18.72	1.296	0.199
diffSTS	−0.60	16.79	−3.31	10.86	0.852	0.397
apeSVQ	24.69	22.58	22.66	15.61	0.468	0.641
biasSVQ	4.83	33.28	−2.91	27.62	1.147	0.254
diffSVQ	−0.31	31.76	−8.03	33.40	1.098	0.275

STS, Self-Talk Scale; SVQ, Self-Verbalization Questionnaire; "ape" denotes average percentage error between self and other; "bias" denotes general direction of disagreement; "diff" refers to difference between total scores. Positive bias and diff scores reflect higher other-reported self-talk.

Additional analyses examining only the friend group indicated that level of closeness was unrelated to the various accuracy scores (all ps > 0.28), that participants who lived with their partner did not differ from those who did not live together on those measures (all ps > 0.55), and that those in dating relationships did not differ in their accuracy scores from those in non-dating relationships (all ps > 0.23). Thus, contrary to our prediction, relationship closeness was unrelated to the degree of self-other self-talk accuracy.

We also analyzed the correlations among the major measures separately for the two groups. For the friend group, self-reported and other-rated scores were highly correlated for both the STS [$r(50) = 0.740$, $p < 0.001$] and the SVQ [$r(50) = 0.727$, $p < 0.001$]. For the stranger group, self-reported and other-rated scores were also significantly correlated for both the STS [$r(34) = 0.538$, $p = 0.001$] and the SVQ [$r(34) = 0.644$, $p < 0.001$]. However, the Fisher r-to-z transformation showed that the correlations did not differ significantly for either the STS ($z = 1.55$, $p = 0.12$) or the SVQ ($z = 0.70$, $p = 0.48$).

In summary, we found no evidence that people who know others well showed greater accuracy in rating their partner's typical self-talk levels than did complete strangers. In fact, friends were less accurate in rating their partner's self-reported self-talk frequency than were strangers. There was also no support for the prediction of greater self-other agreement with the private speech measure compared to the STS. The results suggest that when people rate others who they know on their self-talk levels, the raters may be relying more on their assumptions about their partner's typical self-talk than on behavioral observations of the phenomenon. The results are similar to the findings of Carlson et al. (2013), who found that self-reported internalizing personality characteristics were more accurate than informant reports of those characteristics.

All participants seemed to have assumed that their partners talk to themselves as frequently or infrequently as the participants themselves do. There was a tendency toward greater similarity in self-talk scores among friends than strangers. Thus, the greater inaccuracy shown by the friends might be attributable to their assumption that their partners were more similar to themselves than what the strangers assumed about their randomly assigned partners.

This study focused on the question of the agreement between self- and other-reports of self-talk frequency. Results were similar for both self-talk measures. With respect to the validity of the STS or SVQ, it appears that talking to oneself either silently or aloud is not a behavior that friends can accurately assess. The results suggest that relying on other-reports of self-talk frequency is not an effective way to determine the validity of self-talk measures. It is likely that instances of observing or learning about a friend's self-talk episodes are infrequent; future research examining the nature and extent of friends' knowledge of their partner's self-talk would provide additional insight into the relationship between self- and other-reports of self-talk frequency.

General Discussion

The purpose of this set of studies was to determine the accuracy of STS scores. Results from the first two studies provided good support for the argument that self-reports of typical levels of self-talk frequency correspond well with recent situations where participants reported that self-talk had occurred. In particular, Study 1 results showed that respondents' STS scores were consistent with their reports about specific recent experiences where self-talk occurred. The results from Study 2 showed that frequent self-talkers were more likely to report having talked to themselves when STS-related situations occurred, compared to infrequent self-talkers. Study 3 findings indicated that people who knew each other well were less accurate in their assessments of their partner's typical self-talk frequency than were strangers. The latter findings indicate that there may be reasons to doubt the accuracy of knowledgeable informants' assessments of another's typical self-talk.

Consideration of the STS subscales revealed some interesting patterns and differences. In Study 1, all the STS subscales were significantly correlated with reports of the occurrence of self-talk in recent situations. However, in Study 2, self-reported self-critical and social-assessing self-talk frequency were similar for frequent and infrequent self-talkers. Individuals appear to have less difficulty in accurately judging their typical self-reinforcing self-talk frequency than more negative self-talk. It may be the case that self-reinforcing self-talk is associated with less internal conflict or fewer alternative interpretations compared to self-critical or social-assessing self-talk. It is also the case that some STS items refer explicitly to self-talk about communicating with others (e.g., social-assessing and self-managing items pertaining to what the respondent should do or say). Self-talk that pertains to talking with others may be qualitatively different from self-talk pertaining to purely intrapersonal situations. Research on the phenomenological and social interaction aspects of different kinds of self-talk would be an interesting extension of the current findings.

There is a great deal of research documenting that a lack of insight is characteristic of many mental disorders (Oltmanns and Powers, 2012). The distinction between *ego-dystonic* (personal characteristics that conflict with one's self-image) and *ego-syntonic* (characteristics that are consistent with one's self-image) is also relevant to the question of how people view their self-talk. As Oltmanns and Powers (2012) noted, many mental disorders are ego-dystonic, whereas most personality disorders are ego-syntonic. If people consider talking to themselves as an indication of a mental disorder, they may under-report its frequency. If, however, they view self-talk as a personality characteristic (as seems to be the case based on the present findings), they should be less inclined to under-report its frequency. In the present set of studies, we did not examine participants' beliefs about or perceptions of the phenomenon of self-talk. The results from Study 3 suggest that self-talk is viewed by people as ego-syntonic. Addressing how people view the phenomenon of self-talk (e.g., their stereotypes or assumptions about it) and how those perceptions are related to their self-reported frequency is a promising area for future research.

Another direction for future research would be to manipulate the description of the phenomenon of self-talk when it is being measured. For instance, the instructions for the STS state that "[r]esearchers have determined that all people talk to themselves, at least in some situations or under certain circumstances." The perceived appropriateness of talking to oneself may affect the self-reported nature and frequency of self-talk. We would expect that, if informing respondents that researchers have determined that "mentally disturbed people talk to themselves," significant under-reporting of self-talk frequency would likely occur.

The question of why people show individual differences in self-reports of self-talk frequency is an intriguing one. Early childhood experiences might contribute to such differences (e.g., Brinthaupt and Dove, 2012). Self-reports of self-talk might also reflect respondents' awareness of the situations where self-talk occurs. Individual differences in the motivation to recognize or acknowledge one's self-talk might account for some of the differences in STS scores (see Chin et al., 2012). Future research could help to determine the extent to which self-reports of self-talk reflect actual frequency differences rather than differences due to beliefs or motivations with respect to this inner experience.

In their critique of ESM and questionnaire approaches to studying inner experience, Hurlburt and Heavey (2015, p. 156) note that how people answer the prompts or rate the scale items is likely to be a generalization based on "an unspecified mixture of heuristic (recency, availability, etc.), supposition, confirmation bias, and so on." A more accurate understanding of the STS and the present results would be to note that the STS is a measure of whether people notice talking to themselves and how often they recall doing so upon reflection. The present results refer more to respondents' interpretations of "experience and generalities" (Hurlburt and Heavey, 2015, p. 156) than actual, ongoing experiences of talking to themselves. Of course, the methods used in the present studies were designed to assess the validity of the STS rather than the nature and content of currently ongoing self-talk. Based on the results, researchers who use the STS can be confident that its scores are related to respondents' self-reported instances of talking to themselves across different situations. The extent to which the scores on the STS correspond to actual instances of "pristine" inner experiences remains to be seen.

Taken together, the current findings indicate that general self-reported STS scores are good approximations of specific reports of self-talk. Combined with other research supporting the psychometric properties of the STS, the research reported here provides evidence that this measure of self-talk frequency can be used successfully to study individual differences in the phenomenon of self-talk.

Acknowledgments

References

Back, M. D., Schmukle, S. C., and Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *J. Pers. Soc. Psychol.* 97, 533–548. doi: 10.1037/a0016229

Back, M. D., and Vazire, S. (2012). "Knowing our personality," in *Handbook of Self-Knowledge*, eds S. Vazire and T. D. Wilson (New York, NY: The Guilford Press), 131–156.

Beck, A. T. (1976). *Cognitive Therapy and the Emotional Disorders.* New York: New American Library.

Brinthaupt, T. M., and Dove, C. T. (2012). Differences in self-talk frequency as a function of age, only-child, and imaginary childhood companion status. *J. Res. Pers.* 46, 326–333. doi: 10.1016/j.bbr.2011.03.031

Brinthaupt, T. M., and Erwin, L. J. (1992). "Reporting about the self: issues and implications," in *The Self: Definitional and Methodological Issues*, eds T. M. Brinthaupt and R. P. Lipka (Albany, NY: SUNY Press), 137–171.

Brinthaupt, T. M., Hein, M. B., and Kramer, T. E. (2009). The Self-Talk Scale: development, factor analysis, and validation. *J. Pers. Assess.* 91, 82–92. doi: 10.1080/00223890802484498

Brinthaupt, T. M., and Kang, M. (2014). Many-faceted Rasch calibration: an example using the Self-Talk Scale. *Assessment* 21, 157–165. doi: 10.1177/1073191112446653

Carlson, E. N., Vazire, S., and Oltmanns, T. F. (2013). Self-other knowledge asymmetries in personality pathology. *J. Pers.* 81, 155–170. doi: 10.1111/j.1467-6494.2012.00794.x

Carver, C. S., and Scheier, M. F. (1998). *On the Self-Regulation of Behavior.* New York: Cambridge University Press.

Chin, J., Mrazek, M., and Schooler, J. (2012). "Blind spots to the self: limits in knowledge of mental contents and personal dispositions," in *Handbook of Self-Knowledge*, eds S. Vazire and T. D. Wilson (New York, NY: The Guilford Press), 77–89.

Csikszentmihalyi, M., and Larson, R. (2014). "Validity and reliability of the experience-sampling method," in *Flow and the Foundations of Positive Psychology,* ed. M. Csikszentmihalyi (Berlin: Springer Netherlands), 35–54. doi: 10.1007/978-94-017-9088-8_3

Duncan, R. M., and Cheyne, J. A. (1999). Incidence and functions of self-reported private speech in young adults: a self-verbalization questionnaire. *Can. J. Behav. Sci.* 31, 133–136. doi: 10.1037/h0087081

Duncan, R., and Tarulli, D. (2009). "On the persistence of private speech: empirical and theoretical considerations," in *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation*, eds A. Winsler, C. Fernyhough, and I. Montero (New York: Cambridge University Press), 176–187.

Gatersleben, B., Steg, L., and Vlek, C. (2002). Measurement and determinants of environmentally significant consumer behavior. *Environ. Behav.* 34, 335–362. doi: 10.1177/0013916502034003004

Glass, C. R., and Arnkoff, D. B. (1994). Validity issues in self-statement measures of social phobia and social anxiety. *Behav. Res. Ther.* 32, 255–267. doi: 10.1016/0005-7967(94)90120-1

Hardy, J., Oliver, E., and Tod, D. (2009). "A framework for the study and application of self-talk within sport," in *Advances in Applied Sport Psychology: A Review*, eds S. D. Mellalieu and S. Hanton (London: Routledge), 37–74.

Hurlburt, R. T., and Heavey, C. L. (2015). Investigating pristine inner experience: implications for experience sampling and questionnaires. *Conscious. Cogn.* 31, 148–159. doi: 10.1016/j.concog.2014.11.002

Hurlburt, R. T., Heavey, C. L., and Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Conscious. Cogn.* 22, 1477–1494. doi: 10.1016/j.concog.2013.10.003

Kang, M., Bassett, D. R., Barreira, T. V., Tudor-Locke, C., and Ainsworth, B. E. (2012). Measurement effects of seasonal and monthly variability on pedometer-determined Data. *J. Phys. Act. Health* 9, 336–343.

Kendall, P. C., Howard, B. L., and Hays, R. C. (1989). Self-referent speech and psychopathology: the balance of positive and negative thinking. *Cogn. Ther. Res.* 13, 583–598. doi: 10.1007/BF01176069

Kenrick, D. T., and Funder, D. C. (1988). Profiting from controversy: lessons from the person- situation debate. *Am. Psychol.* 43, 23–34. doi: 10.1037/0003-066X.43.1.23

Khodayarifard, M., Brinthaupt, T. M., Zardkhaneh, S. A., and Fard Azar, G. E. (2014). The psychometric properties of the Self-Talk Scale among Iranian university students. *Psychology* 5, 119–126. doi: 10.4236/psych.2014.52019

Klesges, R. C., Eck, L. H., Mellon, M. W., Fulliton, W., Somes, G. W., and Hanson, C. L. (1990). The accuracy of self-reports of physical activity. *Med. Sci. Sports Exerc.* 22, 690–697. doi: 10.1249/00005768-199010000-00022

Klonsky, E. D., Oltmanns, T. F., and Turkheimer, E. (2002). Informant-reports of personality disorder: relation to self-reports and future research directions. *Clin. Psychol. Sci. Pract.* 9, 300–311. doi: 10.1093/clipsy.9.3.300

Leary, M. R. (2004). *The Curse of the Self: Self-Awareness, Egotism, and the Quality of Human Life.* New York: Oxford University Press.

Magura, S., and Kang, S. (1996). Validity of self-reported drug use in high risk populations: a meta-analytical review. *Subst. Use Misuse* 31, 1131–1153. doi: 10.3109/10826089609063969

McGonigle-Chalmers, M., Slater, H., and Smith, A. (2014). Rethinking private speech in preschoolers: the effects of social presence. *Dev. Psychol.* 50, 829–836. doi: 10.1037/a0033909

Mehl, M. R., Gosling, S. D., and Pennebaker, J. W. (2006). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *J. Pers. Soc. Psychol.* 90, 862–877. doi: 10.1037/0022-3514.90.5.862

Mischel, W., Cantor, N., and Feldman, S. (1996). "Principles of self-regulation: the nature of willpower and self-control," in *Social Psychology: Handbook of Basic Principles*, eds E. T. Higgins and A. W. Kruglanski (New York: The Guilford Press), 329–360.

Moskowitz, D. S. (1986). Comparison of self-reports, reports by knowledgeable informants, and behavioral observation data. *J. Pers.* 54, 294–317. doi: 10.1111/j.1467-6494.1986.tb00396.x

Oltmanns, T. F., and Powers, A. D. (2012). "Knowing our pathology," in *Handbook of Self-Knowledge*, eds S. Vazire and T. D. Wilson (New York, NY: The Guilford Press), 258–273.

Oltmanns, T. F., and Turkheimer, E. (2006). "Perceptions of self and others regarding pathological personality traits," in *Personality and Psychopathology*, eds R. F. Krueger and J. L. Tackett (New York, NY: Guilford Press), 71–111.

Palaniappan, U., Cue, R. I., Payette, H., and Gray-Donald, K. (2003). Implications of day-to-day variability on measurements of usual food and nutrient intakes. *J. Nutr.* 133, 232–235.

Reichl, C., Schneider, J. F., and Spinath, F. M. (2013). Relation of self-talk frequency to loneliness, need to belong, and health in German adults. *Pers. Individ. Dif.* 54, 241–245. doi: 10.1016/j.paid.2012.09.003

Schoeneman, T. J. (1981). Reports of the sources of self-knowledge. *J. Pers.* 49, 284–294. doi: 10.1111/j.1467-6494.1981.tb00937.x

Schriber, R. A., and Robins, R. W. (2012). "Self-knowledge: an individual differences perspective," in *Handbook of Self-Knowledge*, eds S. Vazire and T. D. Wilson (New York, NY: The Guilford Press), 105–127.

Shi, X., Brinthaupt, T. M., and McCree, M. (2015). The relationship of self-talk frequency to communication apprehension and public speaking anxiety. *Pers. Individ. Dif.* 75, 125–129. doi: 10.1016/j.paid.2014.11.023

Shrauger, J. S., and Osberg, T. M. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychol. Bull.* 90, 322–351. doi: 10.1037/0033-2909.90.2.322

South, S. C., Oltmanns, T. F., Johnson, J., and Turkheimer, E. (2011). Level of agreement between self and spouse in the assessment of personality pathology. *Assessment* 18, 217–226. doi: 10.1177/1073191110394772

Srivastava, S. (2012). "Other people as a source of self-knowledge," in *Handbook of Self-Knowledge*, eds S. Vazire and T. D. Wilson (New York, NY: The Guilford Press), 90–104.

Uttl, B., Morin, A., and Hamper, B. (2011). Are inner speech self-report questionnaires reliable and valid? *Procedia – Soc. Behav. Sci.* 30, 1719–1723. doi: 10.1016/j.sbspro.2011.10.332

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *J. Pers. Soc. Psychol.* 98, 281–300. doi: 10.1037/a0017908

Vazire, S., and Mehl, M. (2008). Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *J. Pers. Soc. Psychol.* 95, 1202–1216. doi: 10.1037/a0013314

Vazire, S., and Wilson, T. D. (eds). (2012). *Handbook of Self-Knowledge.* New York: The Guilford Press.

Winsler, A., Fernyhough, C., and Montero, I. (eds). (2009). *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation.* New York: Cambridge University Press.

Winsler, A., and Naglieri, J. (2003). Overt and covert verbal problem solving strategies: developmental trends in use, awareness, and relations with task performance in children aged 5 to 17. *Child Dev.* 74, 659–678. doi: 10.1111/1467-8624.00561

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Self-consciousness concept and assessment in self-report measures

Amanda DaSilveira[1], Mariane L. DeSouza[2] and William B. Gomes[1]*

[1] Laboratory of Experimental Phenomenology and Cognition, Institute of Psychology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [2] Graduate Program in Psychology, Universidade Federal do Espírito Santo, Vitória, Brazil

This study examines how self-consciousness is defined and assessed using self-report questionnaires (Self-Consciousness Scale (SCS), Self-Reflection and Insight Scale, Self-Absorption Scale, Rumination-Reflection Questionnaire, and Philadelphia Mindfulness Scale). Authors of self-report measures suggest that self-consciousness can be distinguished by its private/public aspects, its adaptive/maladaptive applied characteristics, and present/past experiences. We examined these claims in a study using 602 young adults to whom the aforementioned scales were administered. Data were analyzed as follows: (1) correlation analysis to find simple associations between the measures; (2) factorial analysis using Oblimin rotation of total scores provided from the scales; and (3) factorial analysis considering the 102 items of the scales all together. It aimed to clarify relational patterns found in the correlations between SCSs, and to identify possible latent constructs behind these scales. Results support the adaptive/maladaptive aspects of self-consciousness, as well as distinguish to some extent public aspects from private ones. However, some scales that claimed to be theoretically derived from the concept of Private Self-Consciousness correlated with some of its public self-aspects. Overall, our findings suggest that while self-reflection measures tend to tap into past experiences and judged concepts that were already processed by the participants' inner speech and thoughts, the Awareness measure derived from Mindfulness Scale seems to be related to a construct associated with present experiences in which one is aware of without any further judgment or logical/rational symbolization. This sub-scale seems to emphasize the role that present experiences have in self-consciousness, and it is argued that such a concept refers to what has been studied by phenomenology and psychology over more than 100 years: the concept of pre-reflective self-conscious.

Keywords: self-consciousness, self-awareness, self-reflection, mindfulness, self-absorption, self-rumination, self-report measures

Introduction

Throughout the 20th century, Western Psychology has understood self-consciousness as an adaptive personality process that entails the natural human disposition of becoming an object of ones' own consciousness (Duval and Wicklund, 1972; Wiley, 1994). Based on such definition, a number of scales related to self-consciousness have been produced in the recent years (see, for example, Trapnell and Campbell, 1999; Grant et al., 2002; Cardaciotto et al., 2008; McKenzie and Hoyle, 2008). As a consequence, it is assumed that there is a growing interest in empirical

investigations on this construct. However, there is current discussion among cognitive scientists and philosophers on what constitutes self-consciousness (Hurley, 1997; Bermúdez, 1998; Gallup et al., 2002; Légrand, 2007; Gallagher and Zahavi, 2008; Metzinger, 2008; Zahavi, 2010).

Traditionally, the conceptualization behind self-consciousness measures relies on William James' and George Mead's definitions of self-consciousness. To become the object of one's own attention, as suggested by Duval and Wicklund (1972), redirects us to the classical study of James (1890), who proposed that to reflect or think about the self requires that the subject (I) becomes the object (Me) of its own thoughts. From a social approach, Mead (1934) suggests that self-consciousness is the act of adopting the perspective of someone else (You) toward one's own self (I). Currently, studies have argued about the distinction of self-consciousness between the act of reflection and the object of reflection (Düsing, 1997; Zahavi, 2010). Although psychological instruments that claim to measure facets of self-consciousness do not typically address such discussions, this study seeks to investigate which facets of self-consciousness are addressed by self-report measures and if there is consistency among the constructs defined by some scales and questionnaires that have currently been used in psychological studies.

In order to analyze the constructs that are usually addressed by self-consciousness measures, we will organize them into three sections, which correspond to three prominent facets of self-consciousness measured by some instruments: its private/public aspects, adaptive/maladaptive applied characteristics, and the present/past experiences in focus.

Private/Public Aspects of Self-Consciousness

The public and private aspects of self-consciousness have been traditionally investigated and measured since the 1970s, when Fenigstein et al. (1975) developed the SCS. The theory behind it was proposed by Mead (1934) and was further operationalized as the theory of objective self-awareness by Duval and Wicklund (1972). The private and public self-consciousness constructs are distinguished based on the direction of the focus of one's own attention, i.e., either inward (the inner feelings and beliefs one has toward oneself), or outward (the beliefs one has about what other people might think about them). This distinction has been criticized (Wicklund and Gollwitzer, 1987), yet many subsequent researchers have supported the differences between focusing on private and public self-characteristics (Franzoi et al., 1990; Grant, 2001; Eichstaedt and Silvia, 2003). Even so, other studies (Trapnell and Campbell, 1999; Grant et al., 2002) have used the SCS measures and focused specifically on its private aspect. In contrast, McKenzie and Hoyle (2008) discussed the presence of negative public and private aspects of self-consciousness as sustained and inflexible self-focused attention, which is also known as self-absorption.

Adaptive/Maladaptive Aspects of Self-Consciousness

The adaptive and maladaptive aspects of self-consciousness emerged as a research topic mainly in the 1990s, reflecting concerns that the attention toward one's self could be associated with both psychological mindedness and well-being (Trudeau and Reich, 1995), as well as with psychological distress (Ingram, 1990; Thomsen et al., 2013) and negative mood states (Wood et al., 1990). The fact that high levels of self-consciousness could be either associated with psychological well-being as well as psychological distress is usually described in the literature as the paradox of self-consciousness (Trapnell and Campbell, 1999; McKenzie and Hoyle, 2008; Simsek, 2013). In this sense, researchers claimed it became necessary to distinguish between the profits related to being aware of one's thoughts and beliefs (the adaptive side of being self-aware), and counterproductive aspects of self-focus not being able to advance the critical thinking (its maladaptive facet).

According to Ingram (1990), psychological distress occurs when ones self-attention was inflexible, thus self-absorption was the product of disproportional, inappropriate, and excessive focus on self-attention. McKenzie and Hoyle (2008) created the Self-Absorption Scale (SAS), which measures the private and public facets of self-absorption. To other researchers, maladaptive self-attention occurs in the context of self-regulation processes, when discrepancies between one's self-evaluative contents and their standards produce a negative affect and, as a consequence, negative psychopathological states (Buss, 1980; Fleckhammer, 2009). Thus, negative affect which is known to be associated with depression and anxiety generates a neurotic self-attentiveness. This neurotic self-attention was called self-rumination by Trapnell and Campbell (1999), and constituted the basis of their instrument, the Reflection–Rumination Questionnaire (RRQ). Rumination was defined as thoughts that frequently recur and are usually unwelcome. In contrast, self-absorption was, by definition, not only related to thoughts, but to any state in which the focus of all of one's internal processes (affect, cognition, attitudes, motives) is directed to the self excessively and in a sustained manner. Rumination was related to psychopathological traits such as neuroticism, whereas self-absorption appeared to be a more generic characteristic. According to Ingram (1990), it is "difficult to find a psychological disorder that is not characterized by a heightened degree of self-focused attention" (Ingram, 1990, p. 165). Yet, rumination has also been associated with artistic creativity, mainly within musicians (Jones et al., 2014).

Past/Present Aspects of Self-Consciousness

The past/present aspects of self-consciousness refer to the temporal instance that qualifies the self-conscious experience. On one hand, self-consciousness is viewed as a reflexive experience, and, thus, as a synonym for self-reflection. To some researchers (Anderson et al., 1996; Creed and Funder, 1998; Silvia, 1999) self-reflection is considered a dimension of private self-consciousness. In this context, it is related to the activity of inspecting and evaluating one's own thoughts, feelings and behaviors. In fact, the etymology of the word reflection (from Latin, reflexus) indicates to bend back, suggesting that the thought needs to have something to which it is related, like a glance through a past experience, so that it can exist. This is consistent with James' (1890) idea of thought and the aforementioned distinction between "I" and "Me". As such, to reflect or think about the self requires the subject to

become the object of their own thoughts; thus, a reflection must refer to some content, an object that is located in past experience.

On the other hand, self-consciousness can be associated to a present moment of self-experience in which one is aware of their experience without any reflexive judgment attached, which is usually investigated in mindfulness studies. Theory and research on self-consciousness focused on the Eastern traditions, more specifically on the concept of mindfulness, have been having a growing interest in the psychological literature (Kabat-Zinn, 2003; Bishop et al., 2004; Rosch, 2007; Hanley and Garland, 2014). Mindfulness is considered to be different from other conscious states, such as self-concept, schemes, and other constructs related to self-reflection: it is solely related to the quality of the conscious experience in the very moment of its occurrence (Cardaciotto et al., 2008); therefore, it should not be associated with reflexive content (Shear and Jevning, 1999). Bishop et al. (2004) suggest that mindfulness has two main components: a sustained attention to the present moment (Awareness) and an open and acceptant attitude toward the experience (Acceptance). This way, consciousness in the present moment is understood as a continuous process of monitoring internal and external events.

Hence, two approaches to self-consciousness regarding its focus are here distinguished: one is related to a reflective instance and presupposes the content to which self-focus is addressed (which is hereby called the 'past' approach), while the other entails a non-themed consciousness of one's present experience as a whole (which is called the 'present' approach). These two features of self-consciousness suggest two distinguishable epistemological focuses: one is focused on the content and information carried out by thoughts and memory (procedural cognition), and the other is focused on the phenomenal, embodied, and situated cognition.

Research Aims

This present study aims to examine which of the facets (private/public aspects, adaptive/maladaptive applied characteristics, and the present/past experiences of self-consciousness) are being considered in self-consciousness self-report measures. Thus, it was expected that the three prominent facets of self-consciousness (private/public aspects, adaptive/maladaptive applied characteristics, and present/past experiences) would be distinguished by a factor analysis. Moreover, in previous validation and adaptation studies (DaSilveira et al., 2011, 2012a,b) in which five SCSs were applied altogether, the authors received feedback from participants claiming that some items from different instruments appear to be very similar and repetitive; thus, we sought to investigate which items from different scales might reflect the same facets underlying self-consciousness. Scales that measure self-consciousness and its related constructs were chosen for their traditional and frequent use in psychological research or for their recent innovation. Five instruments were selected: (1) the SCS (Revised; Scheier and Carver, 1985); (2) the RRQ (Trapnell and Campbell, 1999); (3) the SRIS – Self-Reflection and Insight Scale (Grant et al., 2002); (4) the SAS (McKenzie and Hoyle, 2008); and

(5) the PHILMS – Philadelphia Mindfulness Scale (Cardaciotto et al., 2008). The scales are described in the following section.

Considering the complex relationships found among constructs measured by these scales, our expectations for this research are the following:

(1) A positive association between the constructs of Private Self-Consciousness (SCS) with Reflection (RRQ), Self-Reflection (SRIS), and Private Self-Absorption;

(2) A moderate positive association between Private Self-Consciousness with Rumination (RRQ); and a negative association with Insight (SRIS);

(3) No associations among the constructs Rumination and Reflection (RRQ), and Self-Reflection and Insight (SRIS) with the public aspects of both Self-consciousness and Self-Absorption;

(4) Positive associations between Public Self-consciousness (SCS) with Social Anxiety, and the public aspect of Self-Absorption;

(5) Positive correlation between the Private Self-Absorption (from SAS) and Rumination (from RRQ), as both claim to measure negative aspects of private self-consciousness;

(6) Inverse association between Rumination (from RRQ) and Insight (from SRIS), considering that the authors (Roberts and Stark, 2008) claim that counter-intuitively high levels of self-reflection may be an impediment to developing insight.

(7) Moderate positive association between the construct Reflection (RRQ) with the constructs Acceptance and Awareness (PHILMS), since reflection disclose an inquisitive activity in self-consciousness.

Materials and Methods

Sample and Procedures

The participants were recruited from the general population in Brazil using a snowball sampling method. An initial sample was asked to recruit people from their environment who would be willing to take part in this study. Six hundred and two individuals were recruited, with a mean age of 25.4 years (SD = 9.63). Although not all participants were in direct contact with the researcher and we lack control of the sampling method, a limitation of snowball form of recruitment, it allowed us to have a cost-efficient number of participants from all Brazilian regions: South 69%, Southeast 20%, Northeast 5%, Central-West 1.2%, North 4%. These Brazilian macroregions are composed of states with similar cultural, economical, historical and social aspects, and they are the most widely referred in Brazil because official information given by the Brazilian Institute for Geography and Statistics uses this system. In addition, 4% of respondents were residing abroad at the time of the survey. Of the participants 36.4% were graduate students, 28.2% undergraduate students, 25, 6% had a university degree, and 5% had completed high school. The rest of the participants (4.8%) did not report their schooling. Of the cohort, 63% were women and 37% men. Questionnaires were answered anonymously, and participants filled out a consent

form before taking part in the study. The study was a within-subjects correlational design, and it was approved by the ethics committee of a Brazilian public University, where this study was conducted.

Measurements

The Self-Consciousness Scale – Revised (SCS-R)

The SCS-R (Scheier and Carver, 1985) is a revised version of the SCS (Fenigstein et al., 1975). The theory behind it was proposed by Mead (1934) and further operationalized as the theory of objective self-awareness by Duval and Wicklund (1972). It understands self-consciousness as the activity of becoming the object of one's own thoughts, and claims to measure three constructs related to self-consciousness. Private Self-Consciousness is related to the inward direction of one's thoughts, whereas Public Self-Consciousness is related to the outward direction of one's thoughts, or the ideas and beliefs one has about the impact of their presence on other people. The third subscale is called Social Anxiety and is considered an enfoldment of the Public Self-Consciousness subscale. The authors defined this construct as a consequence that could emerge from some further reflection upon one's own public self-consciousness. Historically, the SCS was extended and later revised by Buss (1980), Carver and Scheier (1981), Pyszczynsky and Greenberg (1987), and Grant (2001). This Scale continues to be adapted for different populations, like the example of its recent version for use with children (Takishima-Lacasa et al., 2014). It originally consists of 23 items measured on a five-point Likert scale, which were divided into three dimensions: Private Self-Consciousness (nine items, such as "I'm always trying to figure myself out"), Public Self-Consciousness (seven items, such as "I'm concerned about the way I present myself"), and Social Anxiety (six items, such as "I have trouble working when someone is watching me"). The Brazilian version used in this study has 22 items and was adapted by Teixeira and Gomes (1995). It presented a sufficient reliability score ($\alpha = 0.73$ and.89 for test–retest), and confirmed the tri-factor structure in accord with the original scale.

The Reflection–Rumination Questionnaire (RRQ)

The RRQ (Trapnell and Campbell, 1999) was built based on the SCS and aims to measure two forms of the Private Self-Consciousness subscale: Self-reflection and Self-rumination, to which the authors also refer to as "Reflection" and "Rumination." In this scale, Reflection was defined in contrast to self-rumination, as a form of self-attention that infers curiosity and entails a philosophical interest in the self, whereas Rumination is the counterproductive aspect of self-reflection. The RRQ has 24 items that are equally divided into Reflection (for example, "My attitudes and feelings about things fascinate me") and Rumination (for example, "My attention is often focused on aspects of myself I wish I'd stop thinking about"). The Brazilian version used in this study was adapted by Zanon and Teixeira (2006) and presented a good reliability score ($\alpha = 0.87$), in addition to confirming the two-factor structure in accord with the original scale.

The Self-Absorption Scale (SAS)

According to McKenzie and Hoyle (2008), the concept of self-absorption is defined as a sustained and inflexible self-focused attention, which constitutes a psychopathological aspect of self-consciousness. To the authors of the SAS, their concept differs from what is measured by the RRQ, once the Trapnell and Campbell's (1999) measure is said to be an exclusive bifurcation from the Private Self-Consciousness subscale by Scheier and Carver (1985). The special attention given to the distinction between public and private aspects of self-consciousness in the SAS is justified by the authors since such distinction has been useful to detect self-conscious emotions and it could contribute in order to integrate the self-absorption construct to the literature on self-consciousness and psychopathology. The original English version of the SAS consists of 17 items, divided into two dimensions. One dimension is related to the private instance of self-absorption and consists of eight items, such as "I think about myself more than anything else." The other dimension is related to the public instance of self-absorption, and has nine items, such as "I find myself wondering what others think of me even when I don't want to." The Brazilian version of the SAS used in this study was adapted by DaSilveira et al. (2011) and had a good reliability score ($\alpha = 0.83$), with one item excluded from the Private Self-Absorption subscale. Thus, the Brazilian version of this scale has 16 items (seven for Private Self-Absorption and nine for Public Self-Absorption).

The Self-Reflection and Insight Scale (SRIS)

The SRIS (Grant et al., 2002) was originally constructed based on the Private Self-Consciousness dimension from the SCS (Scheier and Carver, 1985), and its authors claim that it is an updated version of its Private Self-Consciousness subscale. The authors added the Insight dimension based on findings that had shown another dimension related to Private Self-Consciousness, called the internal state of awareness (Anderson et al., 1996; Creed and Funder, 1998; Silvia, 1999). This dimension was related to a state of internal understanding that one has toward one's own thoughts, feelings, and behaviors. This definition was also used by Grant et al. (2002) to describe the Insight scale. The SRIS has 12 items for Self-reflection (for example, "I frequently take time to reflect on my thoughts") and eight items for Insight (for example, "I am often aware that I am having a feeling, but I often don't quite know what it is," which is a reversed item). The Brazilian version used in this study was adapted by DaSilveira et al. (2012a) and presented satisfactory reliability, with an $\alpha = 0.90$ for Self-reflection and 0.82 for Insight, respectively, 0.91 and 0.87 in the original scale. The 2-factor structure of the scale was confirmed.

The Philadelphia Mindfulness Scale (PHILMS)

The PHILMS (Cardaciotto et al., 2008) was developed to measure mindfulness as a psychological construct. Concepts as mindfulness and self-consciousness are target of a vast discussion in the literature (Bishop et al., 2004). Its meaning is related to the clarity and non-evaluative fluctuation of the attention toward experience (Kabat-Zinn, 2003), with roots in both Buddhist Psychology and meditation practices. In psychological literature, mindfulness is usually distinguished from conscious states such

as self-concept, schemas, and narratives, which have a reflexive judgment attached to it (Shear and Jevning, 1999). Cognitive scientists have struggled to define mindfulness in a way it can be operationally viable for measurement and training (Van Dam et al., 2010). In this tradition, PHILMS stands out as a scale that operationalised mindfulness as a psychological process, targeting populations unfamiliar with meditation practices. PHILMS accounts for the two main components of mindfulness suggested by Bishop et al. (2004): sustaining attention to the present moment (Awareness), and openness and non-attachment to the experience (Acceptance). More specifically, Awareness is related to a continuous monitoring of internal and external events of experience, in a wider sense than just focal attention. Acceptance is believed to put the PHILMS back to the Buddhist tradition concepts of acceptance and detachment, since it describes the experience of events without judgements and interpretation. But, in order to present the construct in an accessible way to participants that are not familiar to Buddhist practices, the authors reversed the items meant to measure it. The result is a group of 10 items formulated in sentences as "I try to distract myself when I feel unpleasant emotions." The complete PHILMS consists of 20 items measured by a five-point Likert scale that were originally divided into two dimensions: Acceptance (10 items), and Awareness (10 items, such as "When I walk outside, I am aware of smells or how the air feels against my face"). The Brazilian version of the scale used in this study was adapted by DaSilveira et al. (2012b) and presented satisfactory reliability, with confirmed factor structure ($\alpha = 0.81$ for Awareness, and 0.85 for Acceptance, same as the original scale).

Data Analysis

Correlations between scores on the subscales of the five instruments were calculated. The reliability of the measures was calculated using Cronbach's alpha. In order to reveal latent structures behind the five SCSs, a Factorial Analysis with Oblimin rotation was calculated, using both total scores provided by the scales and all scale items. All analyses were conducted using the statistical package program SPSS version 11.

Results

Table 1 presents means, SDs, and internal reliabilities for the measures used in the study. The reliabilities were adequate to good, but some measures clearly had greater internal reliability than others (e.g., Self-reflection vs. Public and Private Self-Consciousness). Although not a main hypothesis, independent sample t-tests examined whether there were any sex differences across the measures. The results showed that women had higher scores than men for self-reflection [SRIS; $t(155) = -2.86$; $p < 0.005$].

Table 2 presents correlations between all study measures. Although the negative correlations between Private Self-Consciousness and Acceptance, Public Self-Consciousness and Insight, Public Self-Consciousness and Acceptance, Social Anxiety and Insight, Social Anxiety and Acceptance, and Private Self-Absorption and Acceptance were statistically significant,

TABLE 1 | Descriptive statistics and subscale reliabilities.

Measure	Mean	SD	Cronbach's α
Private Self-Consciousness [Self-Consciousness Scale - Revised (SCS-R)]	34.18	5.06	0.74
Public Self-Consciousness (SCS-R)	27.22	3.91	0.74
Social Anxiety (SCS-R)	17.96	5.67	0.84
Reflexion [Reflection-Rumination Questionnaire (RRQ)]	45.34	9.49	0.89
Rumination (RRQ)	41.36	9.47	0.88
Self-Reflection [Self-Reflection and Insight Scale (SRIS)]	48.79	8.53	0.90
Insight (SRIS)	26.43	6.08	0.81
Private Self-Absorption [Self-Absorption Scale (SAS)]	13.12	5.20	0.82
Public Self-Absorption (SAS)	25.37	6.93	0.83
Acceptance [Philadelphia Mindfulness Scale (PHILMS)]	29.18	7.10	0.86
Awareness (PHILMS)	36.57	6.11	0.82

they were low in magnitude. Because of the sample size, we hereby highlight correlations of moderate magnitudes (i.e., $r > 0.40$, according to the Dancey and Reidy, 2011, suggested parameters) in **Table 2**.

As expected, there were moderate positive associations between Private Self-Consciousness, Reflection and Self-Reflection. And there were low positive associations between Private Self-Consciousness, Rumination and Private Self-Absorption. Low positive associations were also found between Private Self-Consciousness and the two public dimensions of both Self-Consciousness and Self-Absorption, contrary to theoretical expectations. Social Anxiety also had low associations with both Private and Public Self-Consciousness, although it was only expected to correlate with the later.

In accord with the hypotheses, Rumination had a modest negative association with Insight. A low correlation was also found between Rumination and Private Self-Absorption. Surprisingly, there was a high association between Rumination and the public dimension of self-absorption. In fact, Rumination was correlated with the public dimensions subscales (Public Self-Consciousness and Public Self-Absorption) than the private dimensions subscales (Private Self-Consciousness and Private Self-Absorption).

For the last part of the correlational hypothesis, there were low and negative significant correlations between Rumination and Acceptance. However, the Awareness correlation with Rumination was negative but not significant. In fact, the Awareness subscale had low significant correlations with Private Self-Consciousness, Public Self-Consciousness, Reflection, Insight, and Self-Reflection.

As observed in the correlational analyses, several expected theoretical and empirical hypotheses based on previous studies were maintained. However, several theoretical hypotheses were contradicted, as in the public/private aspects of self-consciousness. Private Self-Consciousness, Public Self-Consciousness, Private Self-Absorption, and Public Self-Absorption did not react as expected in most analyses. On the

TABLE 2 | Correlations for the self-consciousness related subscales.

	PRSC	PBSC	SA	REFL	RUM	SR	INS	PRSA	PBSA	ACC	AWA
PRSC	–	**0.40***	0.13**	**0.64***	0.34*	**0.68***	−0.09	0.33*	0.19*	−0.13**	0.37*
PBSC		–	0.19*	0.16*	0.39*	0.28*	−0.14**	0.16*	**0.40***	−0.19*	0.21*
SA			–	−0.01	**0.42***	−0.01	−0.33*	0.24*	**0.44***	−0.29*	−0.07
REFL				–	0.11	**0.68***	0.03	0.22*	−0.02	0.09	0.27*
RUM					–	0.22*	**−0.52***	**0.47***	**0.60***	**−0.47***	−0.02
SR						–	0.06	0.22*	0.05	−0.05	0.36*
INS							–	**−0.42***	**−0.49***	**0.51***	0.22*
PRSA								–	**0.49***	−0.31*	0.01
PBSA									–	**−0.42***	−0.03
ACC										–	−0.05
AWA											–

Correlation is significant at the 0.01 level (two-tailed).
**Correlation is significant at the 0.05 level (two-tailed).*
Acronyms legend: PRSC, Private Self-Consciousness (SCS-R); PBSC, Public Self-Consciousness (SCS-R); SA, Social Anxiety (SCS-R); REFL, Reflection (RRQ); RUM, Rumination (RRQ); SR, Self-Reflection (SRIS); INS, Insight (SRIS); PRSA, Private Self-Absorption (SAS); PBSA, Public Self-Absorption (SAS); ACC, Acceptance (PHILMS); AWA, Awareness (PHILMS).
All values in bold are correlations higher than 0.40.

other hand, the maladaptive/adaptive distinction involving the Rumination, Private Self-Absorption, and Public Self-Absorption scales was maintained without contradiction.

To clarify the relationship patterns found in the correlations between the self-consciousness measures and identify latent constructs, total scores from the 11 subscales were subjected to a factor analysis, using Oblimin rotation. The Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis, KMO = 0.78. Bartlett's Test of Sphericity χ^2 (55) = 1528.78, $p < 0.001$. Three factors with an Eigenvalue greater than 1 were extracted. **Table 3** shows the loadings for each scale on the relevant factor and the variance explained by the factor. Together, these factors account for 63.3% of the variance. Interestingly,

TABLE 3 | Pattern loadings for the principal axis factoring analysis with Oblimin rotation and Kaiser Normalization.

Measure	F1	F2	F3
Private Self-Consciousness (SCS-R)		**0.82**	0.45
Public Self-Consciousness (SCS-R)			**0.61**
Social Anxiety (SCS-R)	**0.50**		
Reflection (RRQ)		**0.80**	
Rumination (RRQ)	**0.74**		
Self-Reflection (SRIS)		**0.60**	
Insight (SRIS)	**−0.56**		
Private Self-Absorption (SAS)	**0.62**	0.35	
Public Self-Absorption (SAS)	**0.78**		0.32
Acceptance (PHILMS)	**−0.55**		
Awareness (PHILMS)			0.30
Eigenvalue	3.4	2.1	1.0
Variance (%)	31.7	19.5	9.1
Total Variance = 60.4%			
Factor correlation matrix: F1	1.0	0.11	0.10
Factor correlation matrix: F2		1.0	0.27
Factor correlation matrix: F3			1.0

All values in bold are correlations higher than 0.40.

the first rotated factor displayed all measures related to the maladaptive dimensions of self-consciousness (Social Anxiety, Rumination, and Private and Public Self Absorption), plus the variables of the subscales Insight and Acceptance with negative scores. The second factor had high positive loadings for the adaptive characteristics of self-reflection [i.e., Private Self-Consciousness, Reflection (from RRQ) and Self-Reflection (SRIS)]. The third factor stands out public self-consciousness subscale following by the private self-consciousness subscale with lower scores. Interestingly, the subscale of public self-absorption was the third highest scoring in that factor (0.32) very close to the cut-off point of 0.35. The subscale Awareness did not significantly load on any factor. Public Self-Consciousness, which initially loaded on both Factors 2 and 3, appeared closer to F2 in the Factor Plot.

All 102 items from the 11 subscales were subjected to a factor analysis using Oblimin rotation. Twenty-two factors with Eigenvalues greater than 1 and 10 factors with Eigenvalues greater than 1.75 were extracted (O'Connor, 2000). Together, the 22 factors account for 63.4%, while 11 factors account for 50.3% of the variance. According to O'Connor (2000), we used a parallel analysis engine to aid in determining the number of factors to retain (suggested by Patil et al., 2007), and it suggested proceeding a factorial analysis retaining four factors. This procedure aimed to identify how the items would react when forced to fit the smaller number of factors.

The 4-factor solution had a KMO = 0.89, Bartlett's Test of Sphericity χ^2 (5151) = 23932.9, $p < 0.001$. The distribution of items of each Scale according to the factor in which they loaded can be seen on **Table 4**. Since this analysis includes 102 items, the authors chose to display only the factor loadings higher than 0.35. Some items loaded in more than one factor, as noted on **Table 4**. Factor 1 could be called "reflection", since it collected the items related to the adaptive and healthy facet of reflecting upon one's own self. It grouped together all the items from the Reflection and Self-Reflection subscales (from the RRQ and SRIS scales,

TABLE 4 | Suggested four-factor model (pattern loadings) for the 102 items of five Scales (Principal Axis Factoring with Oblimin Rotation and Kaiser Normalization).

Scales Factors	Private and Public Self-Consciousness and Social Anxiety (SCS)			Self-Reflection and Insight (SRIS)			Private and Public Self-Absorption (SAS)			Reflection–Rumination (QRR)			Acceptance and Awareness PHILMS		
	Item	F.L.	Ss	Item	F.L.	Ss	Item	F.L.	Ss	Item	F.L.	Ss	Item	F.L.	Ss
F1 – Reflection	1	(0.65)	PRSC	1	(0.54)	SR	4	(0.37)	PRSA	13	(0.53)	REFL			
(Adaptive and health facet of	4	(0.67)	PRSC	2	(0.55)	SR				14	(0.57)	REFL			
self-consciousness)	8	(0.53)	PRSC	5	(0.58)	SR				15	(0.72)	REFL			
Eigenvalue	12	*(0.45)	PRSC	7	(0.72)	SR				16	(0.50)	REFL			
14.507	14	(0.53)	PRSC	8	(0.71)	SR				17	(0.67)	REFL			
% of Variance	17	(0.46)	PRSC	10	(0.55)	SR				18	(0.68)	REFL			
14.22				12	(0.70)	SR				19	(0.36)	REFL			
Factor correlation with				13	(0.53)	SR				20	(0.58)	REFL			
F2: 0.16				15	(0.67)	SR				21	(0.59)	REFL			
F3: −0.13				16	(0.58)	SR				22	(0.56)	REFL			
F4: −0.38				18	(0.65)	SR				23	(0.66)	REFL			
				19	(0.62)	SR				24	(0.69)	REFL			
F2-Maladaptive facet of	2	(0.37)	PBSC				1	(0.62)	PBSA	1	(0.44)	RUM			
self-consciousness	3	(0.51)	SA				2	(0.46)	PBSA	2	*(0.69)	RUM			
Eigenvalue	5	(0.49)	PBSC				3	(0.53)	PBSA	3	*(0.54)	RUM			
12.09	6	(0.39)	PRSC				7	(0.42)	PBSA	4	(0.57)	RUM			
% of Variance	7	(0.41)	SA				8	(0.44)	PBSA	5	*(0.65)	RUM			
11.86	9	(0.50)	SA				10	(0.54)	PBSA	6	(0.46)	RUM			
Factor correlation with	13	(0.41)	PBSC				13	(0.43)	PBSA	7	(0.60)	RUM			
F3: 0.20	15	(0.40)	SA				16	(0.50)	PBSA	8	(0.59)	RUM			
F4: −0.002	16	(0.34)	PBSC				17	(0.58)	PBSA	9	(0.49)	RUM			
	18	(0.64)	PBSC							10	*(0.39)	RUM			
	22	(0.41)	SA							11	(0.51)	RUM			
										12	*(0.61)	RUM			

(Continued)

TABLE 4 | Continued

Scales Factors	Private and Public Self-Consciousness and Social Anxiety (SCS)			Self-Reflection and Insight (SRIS)			Private and Public Self-Absorption (SAS)			Reflection–Rumination (QRR)			Acceptance and Awareness PHILMS		
	Item	F. L.	Ss	Item	F. L.	Ss	Item	F. L.	Ss	Item	F. L.	Ss	Item	F. L.	Ss
F3 – Avoidance				4	(0.57)	INS	5	(−0.37)	PRSA	2	*(−0.49)	RUM	2	(0.30)	ACC
Eigenvalue				9	(0.36)	INS	6	(−0.41)	PRSA	3	*(−0.44)	RUM	4	(0.57)	ACC
3.93				11	(0.45)	INS	9	(−0.39)	PRSA				6	(0.65)	ACC
% of Variance				14	(−0.56)	INS	11	(−0.47)	PRSA				8	(0.58)	ACC
3.85				17	(−0.61)	INS	15	(−0.49)	PRSA				10	(0.66)	ACC
Factor correlation with				20	*(0.36)	INS							12	(0.68)	ACC
F4: 0.10													14	(0.49)	ACC
													16	(0.63)	ACC
													18	(0.40)	ACC
													20	(0.53)	ACC
F4- awareness of self/	10	(0.42)	PBSC	3	(0.56)	INS				5	*(0.36)	RUM	1	(0.54)	AWA
awareness of experience".	12	*(0.44)	PRSC	6	(0.47)	INS				10	*(0.38)	RUM	3	(0.40)	AWA
(being fully wake or present)	19	(0.45)	PRSC	20	*(0.51)	INS				12	*(0.38)	RUM	5	(0.35)	AWA
Eigenvalue	20	(0.44)	PBSC										7	(0.40)	AWA
2.80	21	(0.45)	PRSC										9	(0.41)	AWA
% of Variance													11	(0.57)	AWA
2.80													13	(0.53)	AWA
													15	(0.41)	AWA
													17	(0.64)	AWA
													19	(0.58)	AWA

Legend:
Item: Number of the item on its correspondent Scale.
FL.: Factor Loadings – values higher than 0.35 are presented between parentheses.
*Indicates items that appear in more than one factor.
Ss: Sub-scale - acronyms legend for each sub-scale: PRSC, Private Self-Consciousness (SCS-R); PBSC, Public Self-Consciousness (SCS-R); SA, Social Anxiety (SCS-R); REFL, Reflection (RRQ); RUM, Rumination (RRQ); SR, Self-Reflection (SRIS); INS, Insight (SRIS); PRSA, Private Self-Absorption (SAS); PBSA, Public Self-Absorption (SAS); ACC, Acceptance (PHILMS); AWA, Awareness (PHILMS).

respectively), as well as items from the Private SCS. Examples of items in Factor 1 are: "I am always trying to figure myself out" (Private Self-Consciousness subscale), all items from the Reflection subscale of the RRQ as "I love analyzing why I do things," and all items from the Self-Reflection subscale of the SRIS, as "It is important for me to evaluate the things that I do." Moreover, one single item of the private aspect of the SAS also belongs to this factor, the item "I think about myself more than anything else."

Factor 2 could be called the maladaptive facet of self-consciousness, since it grouped items that refer to ruminative thoughts and some sense of avoidance toward thoughts and the act of driving one's attention toward themselves, either in private or public situations. It includes items such as "I'm concerned about my style of doing things" (from the Private Self-Consciousness subscale), "I feel anxious when I speak in front of a group" (from the Social Anxiety subscale), "I'm concerned about the way I present myself" (from the Public Self-Consciousness subscale), "I find myself wondering what others think about me even when I don't want to" and "When I start thinking about how others see me, I get all worked up" (both from the Public Self-Absorption subscale), "I tend to 'ruminate' or dwell over things that happen to me for a really long time afterward." and "I often reflect on episodes of my life that I should no longer concern myself with" (both from the RRQ's Rumination subscale).

Factor 3 could be called avoidance, since it gathered all items related to insights and impediments in deliberations one deals toward their own thoughts and behaviors. This factor groups together all reversed items from the Insight subscale, such as "My behavior often puzzles me" and "I often find it difficult to make sense of the way I feel about things," as well as all Private Self-Absorption items, such as "When I have to perform a task, I do not do it as well as I should because my concentration is interrupted with thoughts of myself instead of the task," "My mind never focus on things other than myself for very long." It also has all the items from the Acceptance subscale, such as "When I have a bad memory, I try to distract myself to make it go away" and "I tell myself that I shouldn't feel sad."

Finally, Factor 4 could be called "awareness of self/ awareness of experience," since it not only gathered all the items from the Awareness subscale (such as "I am aware of what thoughts are passing through my mind" and "When I walk outside, I am aware of smells or how the air feels against my face") but it also collects all the items that either describe this present moment of experiences one is having or explicitly contain the verb "aware" in their descriptions, such as "I usually know why I feel the way I do" and "I am usually aware of my thoughts" (both from the SRIS), and "I'm alert to changes in my mood" and "I'm usually aware of my appearance" (both items from the SCS).

Discussion

As for the correlational analyses, Insight scores were negatively correlated with Rumination, which corroborates the theoretical assumption proposed by Roberts and Stark (2008). Similarly, both Acceptance and Insight scores had a modest positive

correlation ($r = 0.51$) and negatively loaded on the same factor in our first Factorial Analysis. In addition, Factor 2 was composed of items that are related to self-consciousness as a self-reflective activity, including the scores of Reflection (RRQ) and Self-Reflection (SRIS) next to the Private Self-Consciousness subscale. Since the Awareness scores did not load on the same factor, we suggest that the Awareness component of mindfulness, as measured by the PHILMS, may not reflect self-reflection. Indeed, the Awareness subscale scores had also presented significant but low correlations with self-reflective measures. The fact that the correlation between Factor 1 and Factor 4 was -0.38 could be used to support the argument that self-reflection is distinct from Awareness. Thus, Awareness appeared to resemble a construct related to the present experience of which one is cognizant without any further judgment nor logical/rational symbolisation.

The Factorial Analysis from the 11 subscale scores provided interesting empirical evidence supporting the maladaptive/adaptive distinction in self-consciousness measures. Factors 1 and 2 have gathered either the counterproductive aspects of self-consciousness or its aspect associated with psychological mindedness and well-being. Note that scores of Public SCS loaded on a separate factor alongside with Private Self-Consciousness. Scores for the Public aspects of Self-Absorption loaded on the same factor but did not reach the minimal loading (i.e., 0.35).

A very similar pattern could be observed in the Factorial Analysis of the 102 items altogether. Items from Factors 1 and 2 reflected the adaptive/maladaptive aspects of self-consciousness. Self-consciousness private aspects predominantly carried out in Factor 1, while the public aspects concentrated in Factor 2. Factor 3 brought a consistent association of one's promptitude for discernment of problems and situations (Insight), combined with one's openness to inner and outer experiences (Acceptance), as well as a sustained and inflexible inward self-focused attention (Private Self-Absorption). It is interesting to note that the Insight items as well as the items from Private Self-Absorption were negative; i.e., they were together in the same factor, but still on the opposite quadrant when compared to the Acceptance items. Such behavior makes theoretical sense, hence Insight and Private Self-Absorption are both constructs related to judging the content of one's thoughts, whereas Acceptance is related to non-judgemental openness to experiences, but, it is described in the PHILMS scale as reversed items; thus, it is also referring to acts that people entail in order to block such judgmental thoughts toward their experiences. Factor 4 offers some evidence for open-mindedness, the receptiveness to the present experience and to new ideas. Thus, in the empirical analyses of the scales, two out of the three prominent facets of self-consciousness were prominent: the adaptive/maladaptive applied characteristics, and the present/past experiences that are focused by the self.

Final Considerations

Self-consciousness, as a construct that is assessed by self-report measures, can be distinguished by its private and public aspects, adaptive/maladaptive applied characteristics, and the present/past experiences that are focused by the self. In this study, we compared these three components by examining

their associations with each other (**Table 3**). The findings suggested that it is possible to argument in favor of a distinction between the public and private aspect of self-consciousness. The projection of public self-consciousness (**Table 3**, third factor) is notorious because most of these scales are explorations of the Private Self-Consciousness. The distinction between adaptive and maladaptive dimensions of self-reflection was sustained by the correlation as well as the factor analyses. Nevertheless, the data presented should also be tested with a neuroticism measure to identify the underlying factor behind such differences. Additionally, in all analyses (correlations and factor analysis), it was observed that the Awareness subscale of the PHILMS was always distinguishable, which suggests that it can be an evidence of pre-reflective aspects of self-consciousness. Moreover, it demonstrates the capacity of human consciousness to establish the condition of here and now, which is the sense of present moment.

When analyzing the factor structure for the total subscale scores, Insight behaved differently than other self-reflective constructs. In the factor analysis of all items, many items from the Insight subscale negatively loaded on the same factor as the maladaptive items. As previously mentioned, such findings corroborated several theoretical expectations (Roberts and Stark, 2008). However, Grant et al. (2002) stated that both subscales, Self-Reflection and Insight, were sub-dimensions of the private self-consciousness construct. Thus, it should be expected that Insight would load in the same factor as Self-Reflection. According to Grant et al. (2002) Insight was a synonym for an internal state of awareness (Anderson et al., 1996; Creed and Funder, 1998; Silvia, 1999). This state has clear theoretical similarities to what the Awareness subscale claims to measure, which suggests a need for further studies to clarify the differences between the Awareness and Insight, and determine which measure accounts for the internal state of awareness, as noted (Creed and Funder, 1998).

In summary, our findings suggest: (1) that in a non-clinical sample, both adaptive/maladaptive and past/present dimensions of self-consciousness were relatively stable structures; and (2) that in spite of variations in the formulation of the scales items, the structural model of Self-Consciousness in James and Mead prevails the subject (I) becomes the object (Me) of its own thoughts. Further studies are needed to confirm that the awareness factor measures what the phenomenology tradition understands as pre-reflective self-consciousness.

References

Anderson, E. M., Bohon, L. M., and Berrigan, L. P. (1996). Factor structure of the private self-consciousness scale. *J. Pers. Assess.* 66, 144–152. doi: 10.1207/s15327752jpa6601_11

Bermúdez, J. (1998). *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., et al. (2004). Mindfulness: a proposed operational definition. *Clin. Psychol. Sci. Pract.* 11, 230–241. doi: 10.1093/clipsy.bph077

Buss, A. H. (1980). *Self-Consciousness and Social Anxiety*. San Francisco, CA: Freeman.

Cardaciotto, L., Herbert, J. D., Forman, E. M., Moitra, E., and Farrow, V. (2008). The assessment of present-moment awareness and acceptance: the Philadelphia mindfulness scale. *Assessment* 15, 204–223. doi: 10.1177/1073191107311467

Carver, C. S., and Scheier, M. F. (1981). *Attention to Self-regulation: A Control-Theory Approach to Human Behavior*. New York, NY: Springer Verlag. doi: 10.1007/978-1-4612-5887-2

Creed, A. T., and Funder, D. C. (1998). The two faces of private self-consciousness: self-report, peer-report, and behavioral correlates. *Euro. J. Personal.* 12, 411–431. doi: 10.1002/(SICI)1099-0984(199811/12)12:6<411::AID-PER317>3.0.CO;2-C

Dancey, C. P., and Reidy, J. (2011). *Statistics Without Maths for Psychology*. Harlow: Pearson Education.

DaSilveira, A. C., DeCastro, T. G., and Gomes, W. B. (2011). Adaptation of the self-absorption scale to Brazilian adults [adaptação da escala de autoabsorção para adultos brasileiros]. *Interação em Psicologia* 15, 5–30.

DaSilveira, A. C., DeCastro, T. G., and Gomes, W. B. (2012a). Self-Reflection and insight scale: new self-consciousness measure adapted and validated to Brazilian adults [escala de autorreflexão e insight: nova medida de autoconsciência adaptada e validada para adultos brasileiros]. *Psico-PUCRS* 43, 155–162.

DaSilveira, A. C., DeCastro, T. G., and Gomes, W. B. (2012b). Adaptation and validation of the Philadelphia mindfulness scale among Brazilian adults [adaptação e validação da escala filadélfia de mindfulness para adultos brasileiros]. *Psico-USF* 17, 215–223.

Düsing, K. (1997). *Selbstbewußtseinmodelle: Moderne Kritiken und Systematische Entwürfe Zur Konkreten Subjektivität*. München: Wilhelm Fink Verlag.

Duval, T. S., and Wicklund, R. A. (1972). *A Theory of Objective Self Awareness*. New York, NY: Academic Press.

Eichstaedt, J., and Silvia, P. J. (2003). Noticing the self: implicit assessment of self-focused attention using word recognition latencies. *Soc. Cogn.* 21, 349–361. doi: 10.1521/soco.21.5.349.28686

Fenigstein, A., Scheier, M. F., and Buss, A. H. (1975). Public and private self-consciousness: assessment and theory. *J. Consult. Clin. Psychol.* 36, 1241–1250. doi: 10.1037/h0076760

Fleckhammer, L. E. (2009). *Insight into the Self-Absorption Paradox: Self-Conscious Ruminative and Reflective Thoughts*. Saarbrücken: VDM Verlag Dr. Müller.

Franzoi, S. L., Davies, M. H., and Markwiese, B. (1990). A motivational explanation for the existence of private self-consciousness differences. *J. Pers.* 58, 641–659. doi: 10.1111/j.1467-6494.1990.tb00248.x

Gallagher, S., and Zahavi, D. (2008). *The Phenomenological Mind*. New York, NY: Routledge.

Gallup, G. G., Anderson, J. L., and Shillito, D. P. (2002). "The mirror test," in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds M. Bekoff, C. Allen, and G. M. Burghardt (Chicago: University of Chicago Press), 325–333.

Grant, A. (2001). Rethinking psychological mindedness: metacognition, self-reflection, and insight. *Behav. Change* 18, 8–17. doi: 10.1375/bech.18.1.8

Grant, A. M., Franklin, J., and Langford, P. (2002). The self-reflection and insight scale: a new measure of private self-consciousness. *Soc. Behav. Personal.* 30, 821–836. doi: 10.2224/sbp.2002.30.8.821

Hanley, A. W., and Garland, E. L. (2014). Dispositional mindfulness co-varies with self reported positive reappraisal. *Pers. Individ. Dif.* 66, 146–152. doi: 10.1016/j.paid.2014.03.014

Hurley, S. (1997). Non-conceptual self-consciousness and agency: perspective and access. *Commun. Cogn.* 30, 207–248.

Ingram, R. E. (1990). Self-focused attention in clinical disorders: review and a conceptual model. *Psychol. Bull.* 107, 156–176. doi: 10.1037/0033-2909.107.2.156

James, W. (1890). *Principles of Psychology*. New York, NY: Dover Publications Inc. doi: 10.1037/11059-000

Jones, M. E., Roy, M. M., and Jay, V. (2014). The relationship between reflective rumination and musical ability. *Psychol. Aesthet. Creat. Arts* 8, 219–226. doi: 10.1037/a0035634

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: past, present, and future. *Clin. Psychol. Sci. Pract.* 10, 144–156. doi: 10.1093/clipsy.bpg016

Légrand, D. (2007). Pre-reflective self- as-subject from experiential and empirical perspectives. *Conscious. Cogn.* 16, 583–599. doi: 10.1016/j.concog.2007.04.002

McKenzie, K., and Hoyle, R. H. (2008). The self-absorption scale: reliability and validity in non-clinical samples. *Pers. Individ. Dif.* 45, 726–731. doi: 10.1016/j.paid.2008.07.020

Mead, G. H. (1962). *Mind, Self and Society. From the standpoint of a Social Behaviorist.* Chicago: University of Chicago Press. (Original published in 1934).

Metzinger, T. (2008). Empirical perspective from the self-model theory of subjectivity: a brief summary with examples. *Prog. Brain Res.* 168, 215–245. doi: 10.1016/S0079-6123(07)68018-2

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP Test. *Behav. Res. Methods Instrum. Comput.* 32, 396–402. doi: 10.3758/BF03200807

Patil, V. H., Singh, S. N., Mishra, S., and Donavan, D. T. (2007). *Parallel Analysis Engine To Aid Determining Number Of Factors To Retain* [Computer Software]. Available at http://ires.ku.edu/~smishra/parallelengine.htm

Pyszczynski, T., and Greenberg, J. (1987). "Toward an integration of cognitive and motivational perspectives on social inference: a biased hypothesis-testing model," in *Advances in Experimental Social Psychology*, ed. L. Berkowitx (San Diego, CA: Academic Press), 20, 297–340. doi: 10.1016/s0065-2601(08)60417-7

Roberts, C., and Stark, P. (2008). Readiness for self-directed change in professional behaviour: factorial validation of the self-reflection and insight scale. *Med. Educ.* 42, 1–10. doi: 10.1111/j.1365-2923.2008.03156.x

Rosch, E. (2007). More than mindfulness: when you have a tiger by the tail, let it eat you. *Psychol. Inq.* 18, 258–264. doi: 10.1080/10478400701598371

Scheier, M. F., and Carver, S. C. (1985). The self-consciousness acale: a revised version for use with general populations. *J. Appl. Soc. Psychol.* 15, 687–699. doi: 10.1111/j.1559-1816.1985.tb02268.x

Shear, J., and Jevning, R. (1999). "Pure consciousness: scientific exploration of meditation techniques," in *The View from Within*, eds F. J. Varela and J. Shear (Thorverton: Imprint Academics), 189–209.

Silvia, P. J. (1999). Explaining personality or explaining variance? A comment on creed and funder (1998). *Euro. J. Personal.* 13, 533–538. doi: 10.1002/(SICI)1099-0984(199911/12)13:6<533::AID-PER354>3.0.CO;2-0

Simsek, O. F. (2013). Self-absorption paradox is not a paradox: illuminating the dark side of self-reflection. *Int. J. Psychol.* 48, 1109–1121. doi: 10.1080/00207594.2013.778414

Takishima-Lacasa, J. Y., Higa-McMillan, C. K., Ebesutani, C., Smith, R. L., and Chorpita, B. F. (2014). Self-consciousness and social anxiety in outh: the revised self-consciousness scales for children. *Psychol. Assess.* 26, 1292–1306. doi: 10.1037/a0037386

Teixeira, M. A. P., and Gomes, W. B. (1995). Self-consciousness scale: a brazilian version. *Psychol. Rep.* 77, 423–427. doi: 10.2466/pr0.1995.77.2.423

Thomsen, D. K., Jensen, A. B., Jensen, T., Mehlsen, M. Y., Pedersen, C. G., and Zachariae, R. (2013). Rumination, reflection and distress: an 8-Month prospective study of colon-cancer patients. *Cogn. Ther. Res.* 37, 1262–1268. doi: 10.1007/s10608-013-9556-x

Trapnell, P. D., and Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: distinguishing rumination from reflection. *J. Pers. Soc. Psychol.* 76, 284–304. doi: 10.1037/0022-3514.76.2.284

Trudeau, K. J., and Reich, R. (1995). Correlates of psychological mindedness. *Pers. Individ. Dif.* 19, 699–704. doi: 10.1016/0191-8869(95)00110-R

Van Dam, N. T., Earleywine, M., and Borders, A. (2010). Measuring mindfulness? An item response theory analysis of the mindful attention awareness scale. *Pers. Individ. Dif.* 49, 805–810. doi: 10.1016/j.paid.2010.07.020

Wicklund, R. A., and Gollwitzer, P. M. (1987). The fallacy of the private-public self-focus distinction. *J. Pers.* 55, 491–523. doi: 10.1111/j.1467-6494.1987.tb00448.x

Wiley, N. (1994). *The Semiotic Self*. Chicago: The University of Chicago Press.

Wood, E., Pressley, M., and Wine, P. (1990). Elaborative interrogation effects on children's learning of factual content. *J. Educ. Psychol.* 82, 741–748. doi: 10.1037/0022-0663.82.4.741

Zahavi, D. (2010). "Inner (time-)consciousness," in *On Time – New Contributions to the Husserlian Phenomenology of Time*, eds D. Lohmar and I. Yamaguchi (New York, NY: Springer Science+Business Media), 319–339. doi: 10.1007/978-90-481-8766-9_16

Zanon, C., and Teixeira, M. A. P. (2006). Adaptation of the rumination and reflection questionnaire (RRQ) to Brazilian university students [adaptação do questionário de ruminação e reflexão (QRR) para estudantes universitários brasileiros]. *Interação em Psicologia* 10, 75–82.

The ARSQ 2.0 reveals age and personality effects on mind-wandering experiences

**B. Alexander Diaz[1,2], Sophie Van Der Sluis[2,3], Jeroen S. Benjamins[4], Diederick Stoffers[1,4],
Richard Hardstone[1,2], Huibert D. Mansvelder[1,2], Eus J. W. Van Someren[1,2,4,5] and
Klaus Linkenkaer-Hansen[1,2]***

[1] Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research, VU University Amsterdam, Amsterdam, Netherlands
[2] Neuroscience Campus Amsterdam, Amsterdam, Netherlands
[3] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, VU University Amsterdam and VU Medical Center Amsterdam, Amsterdam, Netherlands
[4] Department of Sleep and Cognition, Netherlands Institute for Neuroscience, Amsterdam, Netherlands
[5] Department of Medical Psychology, VU University Medical Center, Amsterdam, Netherlands

The human brain frequently generates thoughts and feelings detached from environmental demands. Investigating the rich repertoire of these mind-wandering experiences is challenging, as it depends on introspection and mapping its content requires an unknown number of dimensions. We recently developed a retrospective self-report questionnaire—the Amsterdam Resting-State Questionnaire (ARSQ)—which quantifies mind wandering along seven dimensions: "Discontinuity of Mind," "Theory of Mind," "Self," "Planning," "Sleepiness," "Comfort," and "Somatic Awareness." Here, we show using confirmatory factor analysis that the ARSQ can be simplified by standardizing the number of items per factor and extending it to a 10-dimensional model, adding "Health Concern," "Visual Thought," and "Verbal Thought." We will refer to this extended ARSQ as the "ARSQ 2.0." Testing for effects of age and gender revealed no main effect for gender, yet a moderate and significant negative effect for age on the dimensions of "Self," "Planning," and "Visual Thought." Interestingly, we observed stable and significant test-retest correlations across measurement intervals of 3–32 months except for "Sleepiness" and "Health Concern." To investigate whether this stability could be related to personality traits, we correlated ARSQ scores to proxy measures of Cloninger's Temperament and Character Inventory, revealing multiple significant associations for the trait "Self-Directedness." Other traits correlated to specific ARSQ dimensions, e.g., a negative association between "Harm Avoidance" and "Comfort." Together, our results suggest that the ARSQ 2.0 is a promising instrument for quantitative studies on mind wandering and its relation to other psychological or physiological phenomena.

Keywords: Amsterdam Resting-State Questionnaire (ARSQ), consciousness, mind wandering, personality traits, test-retest reliability

INTRODUCTION

Recent estimates suggest that the human brain engages in mind wandering for approximately half of its waking day, thereby generating thoughts and feelings unrelated to current external demands (Killingsworth and Gilbert, 2010). Despite the ubiquity of mind wandering, research into its nature has since the 1970's (Antrobus, 1968; Antrobus et al., 1970; Singer, 1974) received limited attention until recently reinvigorated (for a review, see Smallwood and Schooler, 2006). The state of wakeful rest—or "resting state"—serves a special role in this context, as it is both frequently employed during functional neuroimaging and may be viewed as a model system to study mind wandering relatively free from external demands.

Most investigations of mind wandering have utilized task-based designs (Antrobus et al., 1966; Christoff et al., 2009; Schooler et al., 2011), enabling the detection of mind wandering episodes as a function of task parameters (e.g., difficulty). Yet, our understanding of the content of mind wandering has remained limited possibly due to lack of established instruments and protocols (for pioneering efforts, see Lehmann et al., 1995; Andrews-Hanna et al., 2010; Delamillieure et al., 2010). In an attempt to capture mind-wandering experiences during the resting state in a standardized fashion, we recently presented the Amsterdam Resting-State Questionnaire (ARSQ) as an efficient self-report tool (Diaz et al., 2013). The ARSQ facilitates a quantitative assessment of thoughts and feelings along several dimensions of mind wandering obtained through factor analysis techniques. This has paved the way for investigating associations between mind-wandering experiences and psychological or physiological variables such as mental health (Diaz et al., 2013), but also gender, age, or personality.

Mind wandering may intuitively seem involuntary and unrestrained in nature. However, recent data from our experiments indicated that ARSQ scores remain significantly correlated between assessments almost 1 h apart (Diaz et al., 2013). This observation raised the question as to what extent mind wandering is state-like (e.g., dependent on situational factors) or trait-like, i.e., partly reflecting stable individual differences. Personality traits could be potential contributors to this stability in mind wandering, considering the DSM-IV definition of a personality disorder and its link to inner experience: *"An enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture"* (American Psychiatric Association, 2000). Cloninger's Temperament and Character Inventory (TCI, Cloninger et al., 1993) is an established instrument to quantify personality traits along the dimensions of a psychobiological model. Crucially, this psychobiological model distinguishes between subconscious "temperaments" and conscious "characters." Temperaments are defined as primarily engaging perceptual memory systems, i.e., automatic responses to stimuli and are categorized as "novelty seeking," "harm avoidance," "reward dependence," and "persistence" (Cloninger et al., 1993). Characters by contrast, are related to different conscious concepts of self and are reliant on declarative memory systems (Cloninger et al., 1993; De Fruyt et al., 2000; Watson and Tharp, 2014), e.g., verbal/visual imagination and symbolic reasoning fall in the categories "self-directedness," "cooperativeness," and "self-transcendence." It is precisely these conscious, self-oriented experiences associated with these character dimensions that we expected to exhibit the strongest overlap with mind wandering, itself defined as conscious thought and feelings unrelated to an external task (Smallwood and Schooler, 2006; McVay and Kane, 2010).

In the present study, we first developed an improved extended 10-dimensional model of resting-state experiences, further increasing practical utility by standardizing the number of items per dimension and allowing for the quantification of more qualitative aspects of mind wandering such as visual and verbal thought. This updated model was subsequently used to test for gender and age specific effects on mind-wandering experience, test-retest stability over time-scales up to 32 months between assessments and, finally, the relationship between ARSQ and personality traits of the psychobiological model.

MATERIALS AND METHODS

PARTICIPANTS

Data were obtained from the Netherlands Sleep Registry (NSR, www.sleepregistry.org), a database aimed at sampling multiple questionnaires in a large cohort comprising the full range from very light to very sound sleepers. Registered participants were invited to complete the Amsterdam Resting-State Questionnaire (ARSQ) and a variety of psychometric instruments (see below) using their home computer. All instruments, including the ARSQ, were administered in Dutch. This yielded two large samples of participants filling out the 50-item ARSQ ($n = 882$, 70% females, age range 19–85, mean age 53.9) as described recently (Diaz et al., 2013), and an extended 54-item version ($n = 562$, 76% females, age range 20–86, mean age 54.4). The study protocols presented

here were approved by the institutional review board of the VU University Medical Centre, Amsterdam, The Netherlands.

ONLINE ARSQ ASSESSMENT PROCEDURE

Participants were first (screen 1) asked to enable and test their PC audio equipment (i.e., turn on speakers or put on headphones) and afterwards proceed to the next screen. The following screen (screen 2) started with the following instruction (translated from Dutch): *"A rest period of five minutes will shortly start. It is important that you relax and remain seated quietly behind your computer with your eyes closed. Try not to fall asleep. Instruction: Once you click on 'next,' a timer will count down from 5 minutes to 0. Once the resting period is over, you will hear a beep. Subsequently you will open your eyes and follow the instructions. Should you be interrupted during the 5 minutes by something or someone, you can open your eyes and click on 'Stop' and subsequently on 'Restart' to newly start the 5 minutes rest session. Now, sit down quietly, click 'next' and immediately close your eyes to start the resting period."* The next screen (screen 3), triggered by initiating the experiment, showed the following text: *"If the resting period has ended you will hear a beep. You can stop the beep by clicking 'Stop'. Afterwards click on "next" to proceed to the questions."* Subsequently (screen 4), the participant was briefly informed about how to rate the questions: *"The 5 minutes of rest are over. Now several statements will follow regarding potential feelings and thoughts you may have experienced during the resting period. Please indicate the extent to which you agree with each statement."*

In order to identify invalid trials, participants in the NSR sample indicated at the end of the questionnaire whether the eyes-closed rest session was interrupted or not, with the option to give a detailed reason. The item order for the extended ARSQ was randomized, except for the last two validation items ("I had my eyes closed" and "I was able to rate the statements"). All statements were scored on a five-point Likert-type scale (1–5) with the labels "Completely Disagree," "Disagree," "Neither Agree nor Disagree," "Agree," and "Completely Agree."

INTERNATIONAL PERSONALITY ITEM POOL

A set of 137 personality items (see Table S1 and http://ipip.ori.org) obtained from the international personality item pool (IPIP, Goldberg, 1999; Goldberg et al., 2006) was included and rated "yes" or "no" by a subset of the participants ($n = 502$, 78% female, mean age 54.6). These items are proxy-measures of Cloninger's Temperament and Character Inventory (TCI, Cloninger et al., 1993), quantifying the temperaments "Novelty Seeking," "Harm Avoidance," "Reward Dependence," "Persistence" and characters "Self-directedness," "Cooperativeness," and "Self-Transcendence." The choice for this set of IPIP items for use in the online test battery of the NSR was primarily motivated by the significant relationship between sleep disorders and temperaments and characters (De Saint Hilaire et al., 2005; An et al., 2012). In addition, the IPIP proxy-measures of Cloninger's TCI exhibit both a high mean correlation of 0.86 with the original scale as well as near identical mean reliability measures (Goldberg, 1999). Finally, free public domain instruments with high validity enable large-scale assessments without being limited by costly licensing fees (Goldberg et al., 2006).

DATA PREPARATION AND ANALYSIS

For the 50-item ARSQ (ARSQ 1.0) measurements, data were obtained from a total of 1445 participants, whereas 993 participants completed the recently extended 54-item ARSQ (ARSQ 2.0). In addition to all items included in the ARSQ 1.0, the ARSQ 2.0 contained the following four statements: "I pictured events," "I pictured places," "I had silent conversations," and "I imagined talking to myself" (Table S2). These items served to extend the list of existing ones querying experiences related to images and words, which are important facets of mind wandering (Schooler and Schreiber, 2004; Buckner and Carroll, 2007; Heavey and Hurlburt, 2008; Delamillieure et al., 2010), enabling the specification of the two additional factors "Visual Thought" and "Verbal Thought." Finally, motivated by applicability in clinical settings, we specified a third factor "Health concern" based on the existing ARSQ statements "I felt pain," "I felt ill," and "I thought about my health." As such, "Health concern" aims to capture variability related to extreme discomfort and preoccupation with health status, often observed in clinical samples (Moffic and Pyakel, 1975).

To keep results compatible with previous data (Diaz et al., 2013), both data sets were filtered based on (1) reported interruption, (2) low motivation (rating "Disagree" or lower), (3) low rated ability to remember thoughts/feelings (rating "Disagree" or lower), (4) reported rating inability (rating below "Agree"), (5) not having eyes closed (rating below "Agree"), and (6) exhibiting extreme responses on the majority of items. This conservative procedure left 882 participants for the ARSQ 1.0 and 562 for the ARSQ 2.0 data set. Data were analyzed using MATLAB 2013a (The Mathworks Inc., Natick, MA), and confirmatory factor analyses (diagonally weighted least squares estimator, unit variance identification, zero mean for the latent variables and unit loading identification) were performed using the Lavaan package (Rosseel, 2012) of R (R Core Team, 2013, Vienna, Austria). Correction for multiple comparisons, where applicable, was performed using false discovery rate (Benjamini and Hochberg, 1995).

RESULTS

AN IMPROVED 10-FACTOR AMSTERDAM RESTING-STATE QUESTIONNAIRE

The dimensional structure of the previously published ARSQ 1.0 (Diaz et al., 2013) harbored an unequal number of items per dimension, thereby introducing a difference in scale-interval between some of the factors. In our attempt to improve upon the efficiency and practicality of the existing ARSQ, we first gathered data from the original ARSQ 1.0 (Table S3) and tested whether a model with a reduced number of items would improve its fit statistics. To this end, the lowest-loading items within each factor were removed, keeping three indicators per factor. The remaining 21 ARSQ items (see **Figure 1**, above dashed line) were then used to specify a similar seven-factor model of resting-state cognition compared to our earlier report (Diaz et al., 2013). This new model fit the data well: $\chi^2_{(168, N=882)} = 989.71$, Root-mean-square error of approximation ($RMSEA$) = 0.074, Comparative Fit Index (CFI) = 0.97. To show that this fit was not dependent on a specific composition of participants, the same confirmatory
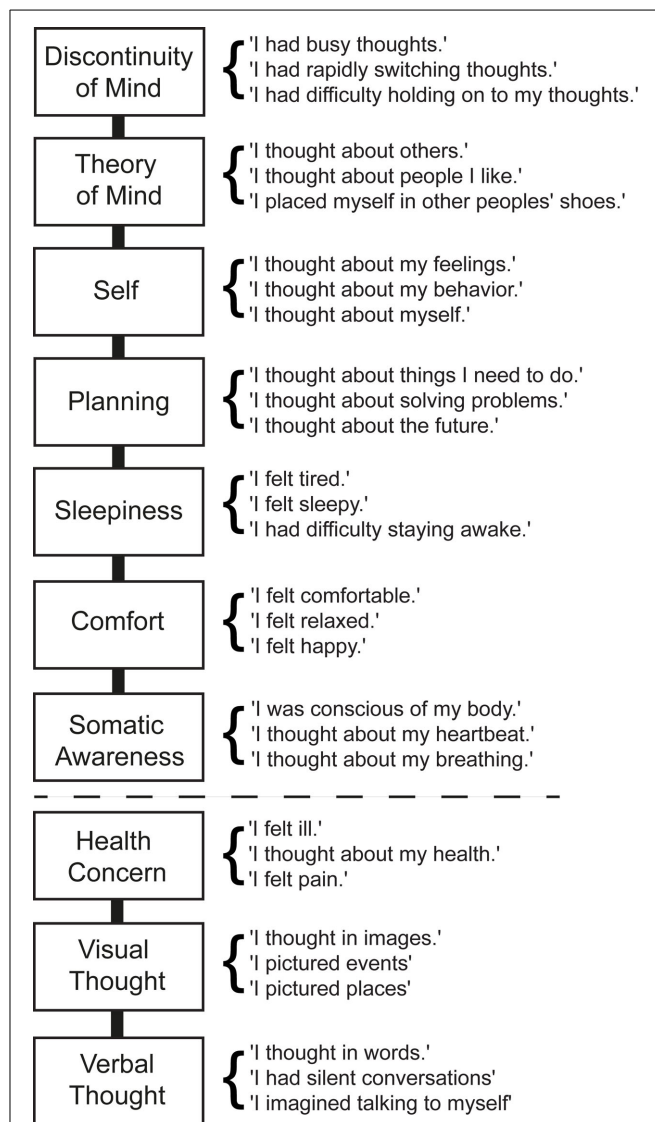

FIGURE 1 | Ten-factor model of mind wandering derived from the Amsterdam Resting-State Questionnaire. The existing model of ARSQ-derived mind-wandering factors (above dashed line, Diaz et al., 2013) can be simplified by reducing the number of items per factor to three. Newly added dimensions tap into verbal and visual thought and "Health Concern."

factor analysis (CFA) was repeated by selecting 600 random cases from the total set (~67% of the data) 1000 times, yielding satisfactory 95% confidence intervals for CFI [0.96, 0.98] and for RMSEA [0.068, 0.081].

Having shown that the abbreviated 7-factor model fits the ARSQ data well, we then proceeded to test an extended 10-factor model (**Figure 1**). The newly specified factors were "Health Concern," "Visual Thought," and "Verbal Thought," each with three items, resulting in 30 items in the model. The "Health Concern" factor includes items already present in the original ARSQ 1.0, whereas the factors "Visual Thought" and "Verbal Thought" demanded the addition of 4 new items (**Figure 1**),

resulting in the ARSQ 2.0. This 10-factor model showed acceptable fit [$\chi^2_{(360,\ N=562)} = 1887.1$, $RMSEA = 0.087$, $CFI = 0.92$] and to show that sample composition was unlikely to have affected the results, the confirmatory factor analysis was replicated 1000 times, each time drawing 420 cases at random (75% of the data, to maintain robust results given the smaller sample size). The resulting 95% confidence intervals were satisfactory for both CFI [0.91, 0.93] and RMSEA [0.079, 0.089].

After having tested the newly specified model, we computed scores for each factor based on the mean of the raw ARSQ 2.0 responses of the respective factor items. This procedure favors straightforward applications, as it avoids estimation of individual factor scores per new sample. The averaging of the scores furthermore produces scores on the same [1, 5] interval as the original ARSQ 1.0 statements, aiding interpretation of absolute values. Correlating the estimated ARSQ 2.0 factor scores to the mean scores (**Table 1**) suggests that compared to our original 7-factor model, reduction of the number of items per factor results in a very good approximation of the estimated factor scores by the mean scores. Notably, the factor "Somatic Awareness" benefits the most from the removal of relatively low-loading items. While using mean scores (as opposed to factor scores) results in lowered factorial correlations, **Figure 2B** shows that these remain significant and in the same direction as the estimated factorial correlations from the CFA model. The strength and significance of the estimated correlations among the factors is further visualized in **Figure 3**.

Finally, to account for potential age and gender effects, we tested the effect of the fixed factor gender on the 10 mean scores ($n = 562$) using multivariate ANOVA, taking age as covariate. The results indicated no overall effect of gender [$F_{(10,\ 541)} = 1.61$, $p = 0.10$] on mind wandering. The only significant difference in mean scores between the genders was found for "Verbal Thought" [$F_{(10,\ 550)} = 4.23$, $p = 0.04$], with women indicating to

have experienced slightly more verbal thoughts ($M = 3.1$, $SD = 0.86$) compared to men ($M = 2.85$, $SD = 0.86$).

However, there was a significant effect of the covariate age [$F_{(10,\ 541)} = 6.51$, $p < 0.001$]. A subsequent correlation analysis between age and the mean scores of the 10-factors showed significant negative correlations for both "Discontinuity of Mind" [$r_{(560)} = -0.20$, $p < 0.001$], "Self," [$r_{(560)} = -0.18$, $p < 0.001$], "Planning," [$r_{(560)} = -0.27$, $p < 0.001$], "Visual Thought," [$r_{(560)} = -0.20$, $p < 0.001$], and "Verbal Thought," [$r_{(560)} = -0.15$, $p < 0.001$]. These results suggest that people experience less verbal and visual thought, less planning and fewer thoughts about themselves with increasing age. To investigate this further, we tested for an age effect on the single items "I felt nothing" and "I thought about nothing," which are part of the full ARSQ 1.0 and ARSQ 2.0 albeit not part of any factor solution. Interestingly, we indeed identified a small yet significant increase in the response to these items with age [both $r_{(560)} = 0.16$, $p < 0.001$].

STABILITY OF ARSQ FACTORS OVER TIME

Recent reports have shown that ARSQ scores are quite stable over a short interval of 45 min (Diaz et al., 2013). This raises the question to what extent ARSQ responses are stable over longer durations. A subset of participants ($n = 216$) filled out both the ARSQ 1.0 and ARSQ 2.0 with a median interval between both measurements of 16.6 months (range: 2.9–31.5 months). We tested the equality of test-retest correlations over time by dividing the total sample ($n = 216$) into 3 equal subgroups (G1: $n = 72$, time between administrations 2.9–14.2 months; G2: $n = 72$, time between administrations 14.3–24.6 months; G3: $n = 72$, time between administrations 24.7–31.5 months). Because the original ARSQ 1.0 lacked the items to form the factors "Visual Thought" and "Verbal Thought," only the single items "I thought in images" and "I thought in words" were used as predictor variables. We observed test-retest correlations of mean scores ranging between 0.34 and 0.54 for the 10 dimensions of the ARSQ (**Table 2**). This suggests that there is a stable component to mind wandering, except for "Sleepiness" and "Health Concern," which are indeed expected to fluctuate on time scales of days or months as part of the normal variation in the amount of sleep obtained or health status.
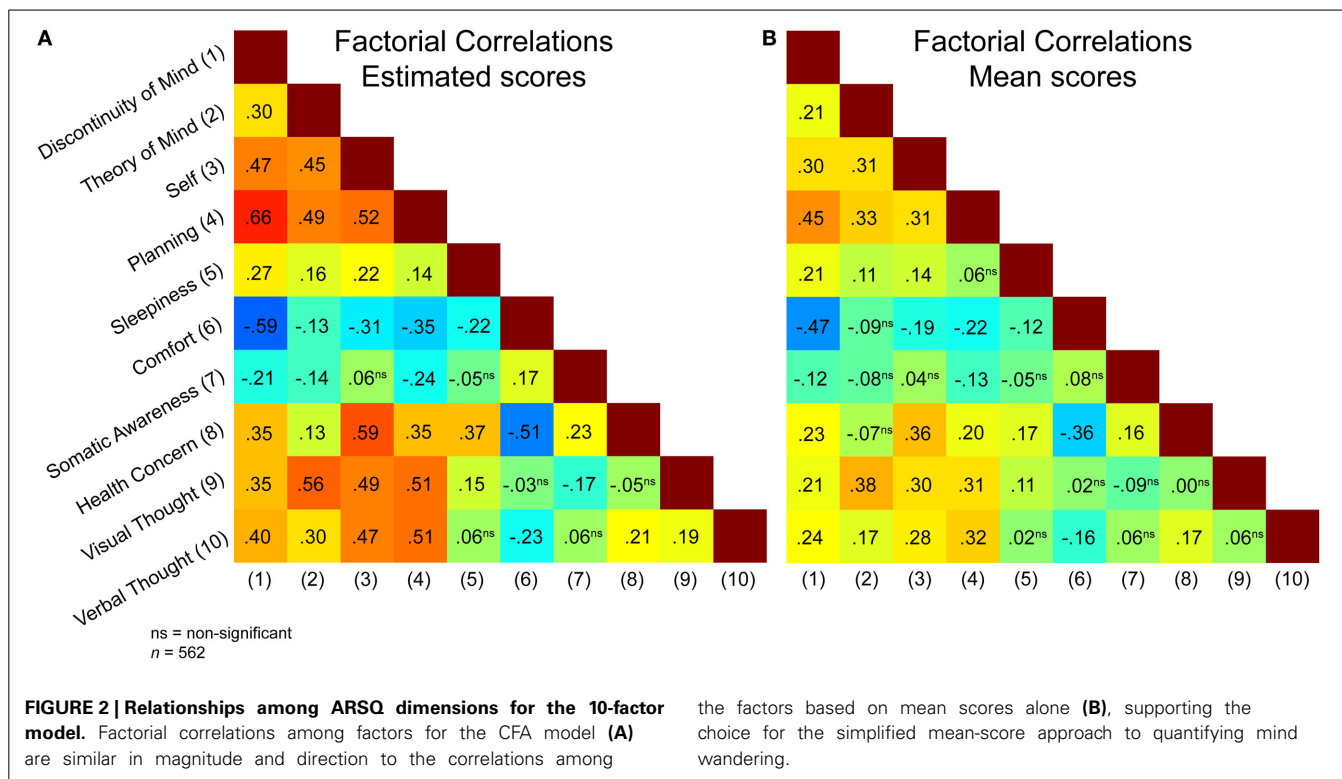
MIND WANDERING AND PERSONALITY TRAITS

To address the question whether the stability in ARSQ factors is related to personality, mean scores for the ARSQ 2.0 sample were correlated with the 3 main character and 4 main temperament dimensions obtained from the personality inventory we administered (see Materials and Methods). As predicted by the psychobiological model (Cloninger et al., 1993; De Fruyt et al., 2000), the character dimensions—which are more dependent on conscious, episodic memory-driven experiences—appeared to be correlated to a larger set of ARSQ factors than the temperaments, which are theorized to reflect more unconscious tendencies (**Table 3**). Notably, "Self-Directedness," an indicator of how well an individual can adapt, control and regulate behavior in response to situational changes and in line with personal goals, appears significantly correlated with most ARSQ factors. For instance, more

---

Table 1 | Correlation between estimated CFA factor scores and scores derived from averaging the item responses within each factor.

	ARSQ 1.0 Correlations 7-factor model[a] n = 813, $p_{corrected} < 0.001$	ARSQ 2.0 Correlations 10-factor model n = 562, $p_{corrected} < 0.001$
Discontinuity of mind	0.95	0.91
Theory of mind	0.98	0.97
Self	0.88	0.90
Planning	0.96	0.90
Sleepiness	0.96	0.96
Comfort	0.94	0.91
Somatic awareness	0.66	0.94
Health concern	–	0.90
Visual thought	–	0.87
Verbal thought	–	0.92

The Pearson correlation coefficients indicate that the mean scores are good predictors of the estimated factor scores also in the ARSQ 2.0 based 10-factor model.

[a]*From (Diaz et al., 2013), p-values false discovery rate corrected.*

FIGURE 2 | Relationships among ARSQ dimensions for the 10-factor model. Factorial correlations among factors for the CFA model **(A)** are similar in magnitude and direction to the correlations among the factors based on mean scores alone **(B)**, supporting the choice for the simplified mean-score approach to quantifying mind wandering.
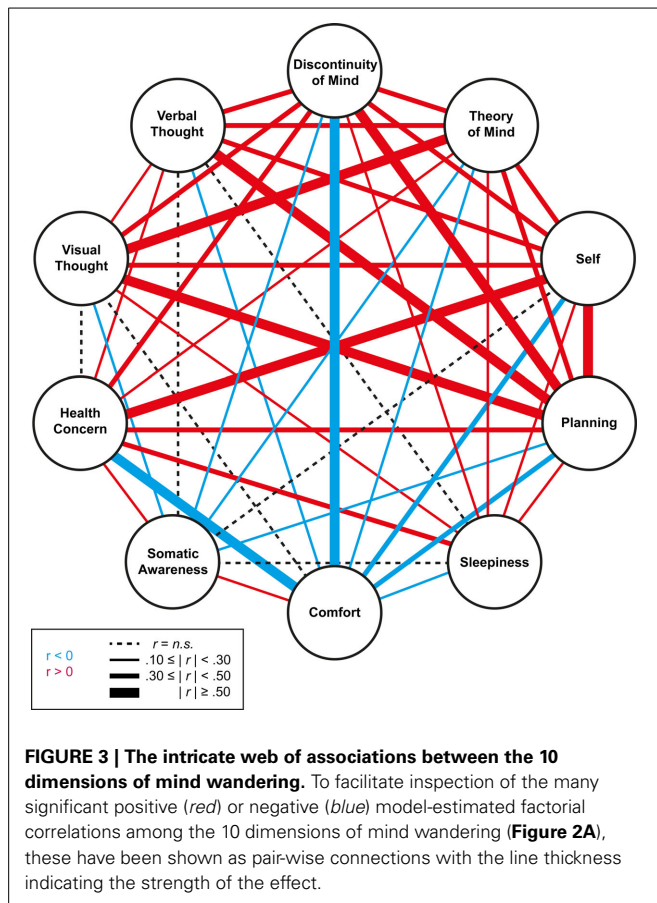
self-directedness is associated with more control over thoughts (i.e., lower scores on "Discontinuity of Mind") and higher scores on Comfort. By contrast, "Self-Transcendence," associated with spiritual traits (e.g., engaging in prayer or meditation), appeared to be largely independent of ARSQ factors. Contrary to expectation, the trait "Cooperativeness," which should reflect individual differences in identification with, and acceptance of others (Cloninger et al., 1993), showed only few significant (and small) correlations with ARSQ factors and did not correlate with "Theory of Mind." On the other hand, "Theory of Mind" was significantly related to "Reward Dependence," viewed as a heritable bias toward seeking approval of and attachment to others (Cloninger et al., 1993). Other temperaments associated with ARSQ factors include "Harm Avoidance," which has been related to neuroticism and fear of uncertainty (De Fruyt et al., 2006) and showed a negative correlation with Comfort and a positive correlation with Health Concern. By contrast, "Novelty Seeking," assumed to reflect impulsivity, showed a positive correlation with Comfort and a negative correlation with Health Concern. Taken together, several of the significant correlations appeared plausible in direction albeit small in magnitude.

To test for age and gender effects on the character and temperament traits, multivariate ANOVA was performed, with gender as fixed-factor and age as covariate. Both the overall main effect of gender [$F_{(7, 492)} = 8.61$, $p = 0.001$] and the covariate age [$F_{(7, 492)} = 6.45$, $p = 0.001$] were significant. Closer inspection of the gender effect revealed higher average "Self-Transcendence" [$F_{(1, 498)} = 14.23$, $p = 0.001$], "Harm Avoidance" [$F_{(1, 498)} = 6.81$, $p = 0.01$], and "Reward Dependence" [$F_{(1, 498)} = 14.18$,

$p = 0.001$] for women [$M = (7.02, 7.78, 12.95)$, $SD = (1.58, 2.92, 2.79)$] compared to men [$M = (6.52, 7.00, 11.81)$, $SD = (1.61, 3.16, 3.05)$], respectively. Finally, a correlation analysis revealed that age had a significant, but low association with self-directedness [$r_{(500)} = 0.11$, $p < 0.01$], cooperativeness [$r_{(500)} = 0.15$, $p < 0.01$], self-transcendence [$r_{(500)} = 0.14$, $p < 0.01$], and novelty-seeking [$r_{(500)} = 0.12$, $p < 0.01$]. These small correlations are in line with the expectation that personality traits change little with age (McCrae and Costa, 1994).

DISCUSSION

The past decade has witnessed a markedly increasing interest in resting-state brain activity, mind wandering and their putative associations (Smallwood and Schooler, 2006; Buckner and Carroll, 2007; Raichle, 2011; Fell, 2013). The resting-state condition also features prominently in clinical research, because of the ease with which patients can perform the task (Linkenkaer-Hansen et al., 2005; Andrews-Hanna et al., 2007; Stoffers et al., 2007; Greicius, 2008; Montez et al., 2009). In spite of the progress that has been achieved on the neuroimaging side, however, efforts toward efficient assessment and quantification of the subjective dimension of mind wandering have been limited (Delamillieure et al., 2010). In our view, integration of standard neuroimaging methodology and the assessment of subjective experience is necessary in order to gain a more complete understanding of mind wandering in general and the functional significance of resting-state brain activity in particular. This formed the driving force behind the development of the ARSQ, a time-efficient and informative tool for resting-state and mind-wandering research (Diaz et al., 2013).

FIGURE 3 | The intricate web of associations between the 10 dimensions of mind wandering. To facilitate inspection of the many significant positive (*red*) or negative (*blue*) model-estimated factorial correlations among the 10 dimensions of mind wandering (**Figure 2A**), these have been shown as pair-wise connections with the line thickness indicating the strength of the effect.

Table 2 | Test-retest correlations between first (ARSQ 1.0) and second (ARSQ 2.0) assessments based on mean scores.

	r_{G1}	r_{G2}	r_{G3}	$r_{averaged}$
Discontinuity of mind	0.55	0.50	0.56	0.54
Theory of mind	0.31	0.33	0.29	0.31
Self	0.44	0.41	0.45	0.43
Planning	0.36	0.13	0.28	0.26
Sleepiness	0.12	0.30	0.57	0.34*
Comfort	0.36	0.56	0.49	0.47
Somatic awareness	0.55	0.28	0.45	0.43
Health concern	0.67	0.45	0.26	0.47*
Visual thought[1]	0.37	0.23	0.31	0.31
Verbal thought[2]	0.35	0.12	0.35	0.27

Testing for equal correlations over three time periods (G1: 2.9–14.2 months; G2: 14.3–24.6 months; G3: 24.7–31.5 months, n = 72 for each group) revealed significant differences only for "Sleepiness" and "Health Concern" (asterisks). This suggests that all other shown factors were stable across the three time intervals as indicated by their averaged correlation.

*n = 216, [1] Single item "I thought in images," [2] Single item "I thought in words," *Correlation showing significant (false discovery rate corrected) differences over groups.*

The current study builds on our earlier efforts, yielding a model of mind wandering with more factors, acknowledging the earlier underrepresentation of important factors such as imagery and inner speech (Schooler and Schreiber, 2004; Heavey and Hurlburt, 2008; Delamillieure et al., 2010) and a standardized set of items per factor in order to avoid differences in the discreetness of the underlying scale. This updated model performs well on a theoretical basis (i.e., improved fit statistics for CFI) and is well approximated by mean item scores. Finally, the addition of a separate factor for "Health Concern," provides an extra estimator of the experience or concern of being ill, often associated with for example patients suffering from depression or chronic pain (Moffic and Pyakel, 1975; Sarnthein et al., 2006).

Our previous results have shown that ARSQ mean scores retain high test-retest correlations over 45 min. Interestingly, even at (much) longer time scales spanning several months between assessments, most correlations remain significant, albeit weaker. Only the factors "Sleepiness" and "Health Concern" did not exhibit stability over time, which is understandable as health status can naturally change over time and sleepiness may be a highly dynamic state (i.e., affected by a restless night, time of day, etc.). To further identify potential contributors to the observed stability in mind wandering, we correlated the ARSQ mean scores to personality measures. We obtained scores on a subset of items closely related to Cloninger's Temperament and Character Inventory (Cloninger et al., 1993; De Fruyt et al.,

2000; Goldberg et al., 2006), measuring three characters, closely related to conscious concept-driven processing and four temperaments, measuring largely unconscious behavioral tendencies. Considering the prominent role of conscious experience in both mind wandering and Cloninger's definition of character traits, we expected significant associations between the ARSQ factors and the three character dimensions. Our results supported this hypothesis partially. The character trait Self-Directedness correlated significantly with most ARSQ factors, and interestingly all significant correlations except for the relation with "Comfort," proved negative. This may suggest that a higher disposition on self-directedness, defined as the individual ability to govern behavior according to situational demands and in line with personal motivators (Cloninger et al., 1993; Watson and Tharp, 2014), is related to *less* mind wandering and *higher* ratings of comfort. Still, the other character traits exhibited far fewer correlations with the ARSQ factors. For example, the character trait Cooperativeness did not correlate with "Theory of Mind," which appears counter-intuitive, as this trait theoretically measures the individual ability to identify with and accept other people—descriptions that fit well with the concept of theory of mind. The temperaments on the other hand, showed specific correlations with up to three ARSQ factors only (**Table 3**). The ARSQ factor "Comfort" appears to be most readily associated with temperaments, adding to face validity via a negative correlation with the neuroticism-related trait Harm Avoidance and a positive relation with Persistence, measuring self-confidence and perseverance.

Overall, our results suggest that the traits of the psychobiological model explain little in terms of variability in ARSQ factors. The possibility exists that other personality inventories, such as the well-known NEO-PI-R (Costa and McCrae, 1985), which measures the "Big Five" (i.e., neuroticism, extraversion, conscientiousness, agreeableness and intellect) would

Table 3 | Correlation matrix of ARSQ 2.0 factors and IPIP personality dimensions based on Cloninger's psychobiological model (Cloninger et al., 1993; Goldberg et al., 2006**).**

	Self-directedness	Cooperativeness	Self-transcendence	Novelty seeking	Harm avoidance	Reward dependence	Persistence
Discontinuity of mind	−0.26*	−0.16*	−0.11*	−0.17*	0.16*	0.03	−0.09
Theory of mind	−0.06	0.03	−0.04	−0.02	0.04	0.15*	−0.01
Self	−0.20*	−0.11*	−0.02	0.01	0.04	0.11*	−0.12*
Planning	−0.18*	−0.16*	−0.07	−0.04	0.07	−0.01	0.00
Sleepiness	−0.12*	−0.06	0.01	−0.05	0.04	0.03	−0.09
Comfort	0.30*	0.13*	0.07	0.22*	−0.32*	0.02	0.20*
Somatic awareness	0.03	0.08	0.06	0.05	−0.01	0.09	0.02
Health concern	−0.18*	0.01	0.01	−0.13*	0.20*	0.02	−0.07
Visual thought	−0.08	−0.12*	−0.08	0.01	0.01	0.03	−0.01
Verbal thought	−0.12*	0.01	0.01	−0.01	0.05	0.07	−0.04

*$p < 0.01$, all p-values corrected using false discovery rate, n = 502.

reveal stronger correlations. Still, previous studies have revealed relatively strong correlations between the Big Five and the dimensions of Cloninger's Temperament and Character inventory (De Fruyt et al., 2006). The search to disentangle stable trait components from more dynamic states and their interaction is an active field of research. Rigorous tracking of mind wandering over longer periods, comparable to "experience sampling" (Zelenski and Larsen, 2000), in combination with advanced statistical methodology such as CFA/structural equation modeling (Steyer et al., 1999) and novel theoretical frameworks, such as conceptualizing traits as state density functions (Fleeson, 2001), may yield a more detailed picture of the state-trait interactions in relation to mind wandering. Current findings suggest that at least part of the ARSQ retest correlations may be governed by "trait-like" components, possibly genetic in nature, and mind-wandering research may therefore benefit from these novel approaches. Alternatively, the (subset of) TCI items we have utilized in this study may be too heterogeneous, i.e., measuring too many distinct facets, preventing clear-cut correlations with the ARSQ. However, inspection of the individual items used (Table S1) for the character "Cooperativeness" and temperament "Reward Dependence" strongly suggests they aim to capture an individual's disposition towards other people, hence they appear congruent with the "Theory of Mind" items of the ARSQ. Perhaps then, there is a discrepancy between how people generally rate themselves, e.g., in personality inventories and their actual experiences during mind wandering, which may be heavily self-referential as suggested by the high factorial correlations for the ARSQ factor "Self" (**Figure 2**).

Finally, we investigated whether age and gender were associated with mind-wandering experience and personality traits. Although both gender and age effects on personality were in line with earlier reports (McCrae and Costa, 1994), only age appeared to have a significant main negative effect on ARSQ dimensions "Planning," "Self," and "Visual Thought." These findings are supported by earlier reports of diminished "current concerns" with age (Klinger, 1999; McVay and Kane, 2010) and could allow future studies to confirm whether age-related variation in neuroimaging measurements can be related to mind wandering.

To conclude, the here presented extended ARSQ and associated 10-factor model improves on the original ARSQ by introducing important additional dimensions. We propose to refer to this revision as "ARSQ 2.0"[1] to distinguish it from the previous version (Diaz et al., 2013). We remain confident that a tool such as the ARSQ will help further develop the "neuroscience of mind wandering" (Gruberger et al., 2011) and help bridge the gap between its behavioral and subjective dimensions.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00271/abstract

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, 4th Edn, Text Revision (DSM-IV-TR).* American Psychiatric Association. Available online at: http://www.citeulike.org/user/caelumluna/article/1140498 [Accessed December 2, 2013]. doi: 10.1176/appi.books.9780890423349

An, H., Park, J., Jang, E.-S., and Chung, S. (2012). The impact of temperament and character on the efficacy of nonpharmacologic treatment of primary insomnia. *Compr. Psychiatry* 53, 201–207. doi: 10.1016/j.comppsych.2011.02.008

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., and Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.* 104, 322. doi: 10.1152/jn.00830.2009

---

[1] The ARSQ is available in multiple translations on request.

Andrews-Hanna, J. R., Snyder, A. Z., Vincent, J. L., Lustig, C., Head, D., Raichle, M. E., et al. (2007). Disruption of large-scale brain systems in advanced aging. *Neuron* 56, 924–935. doi: 10.1016/j.neuron.2007.10.038

Antrobus, J. S. (1968). Information theory and stimulus-independent thought. *Br. J. Psychol.* 59, 423–430. doi: 10.1111/j.2044-8295.1968.tb01157.x

Antrobus, J. S., Singer, J. L., Goldstein, S., and Fortgang, M. (1970). Section of psychology: mindwandering and cognitive structure. *Trans. N.Y. Acad. Sci.* 32, 242–252. doi: 10.1111/j.2164-0947.1970.tb02056.x

Antrobus, J. S., Singer, J. L., and Greenberg, S. (1966). Studies in the stream of consciousness: experimental enhancement and suppression of spontaneous cognitive processes. *Percept. Mot. Skills* 23, 399–417. doi: 10.2466/pms.1966.23.2.399

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.

Buckner, R. L., and Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn Sci* 11, 49–57. doi: 10.1016/j.tics.2006.11.004

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., and Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci.* 106, 8719–8724. doi: 10.1073/pnas.0900234106

Cloninger, C., Svrakic, D. M., and Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Arch. Gen. Psychiatry* 50, 975–990. doi: 10.1001/archpsyc.1993.01820240059008

Costa, P. T. Jr., and McCrae, R. R. (1985). *The NEO Personality Inventory Manual.* Odessa, FL: Psychological Assessment Resources.

De Fruyt, F., De Clercq, B. J., van de Wiele, L., and Van Heeringen, C. (2006). The validity of Cloninger's psychobiological model versus the five-factor model to predict DSM-IV personality disorders in a heterogeneous psychiatric sample: domain facet and residualized facet descriptions. *J. Pers.* 74, 479–510. doi: 10.1111/j.1467-6494.2006.00382.x

De Fruyt, F., van de Wiele, L., and Van Heeringen, C. (2000). Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Pers. Individ. Dif.* 29, 441–452. doi: 10.1016/S0191-8869(99)00204-4

Delamillieure, P., Doucet, G., Mazoyer, B., Turbelin, M. R., Delcroix, N., Mellet, E., et al. (2010). The resting state questionnaire: an introspective questionnaire for evaluation of inner experience during the conscious resting state. *Brain Res. Bull.* 81, 565–573. doi: 10.1016/j.brainresbull.2009.11.014

De Saint Hilaire, Z., Straub, J., and Pelissolo, A. (2005). Temperament and character in primary insomnia. *Eur. Psychiatry* 20, 188–192. doi: 10.1016/j.eurpsy.2004.04.009

Diaz, B. A., Van Der Sluis, S., Moens, S., Benjamins, J. S., Migliorati, F., Stoffers, D., et al. (2013). The Amsterdam Resting-State Questionnaire reveals multiple phenotypes of resting-State cognition. *Front. Hum. Neurosci.* 7:446. doi: 10.3389/fnhum.2013.00446

Fell, J. (2013). Unraveling inner experiences during resting state. *Front. Hum. Neurosci.* 7:409. doi: 10.3389/fnhum.2013.00409

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: traits as density distributions of states. *J. Pers. Soc. Psychol.* 80, 1011–1027. doi: 10.1037/0022-3514.80.6.1011

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personal. Psychol. Eur.* 7, 7–28.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public-domain personality measures. *J. Res. Pers.* 40, 84–96. doi: 10.1016/j.jrp.2005.08.007

Greicius, M. D. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Curr. Opin. Neurol.* 21, 424–430. doi: 10.1097/WCO.0b013e328306f2c5

Gruberger, M., Ben-Simon, E., Levkovitz, Y., Zangen, A., and Hendler, T. (2011). Towards a neuroscience of mind-wandering. *Front. Hum. Neurosci.* 5:56. doi: 10.3389/fnhum.2011.00056

Heavey, C. L., and Hurlburt, R. T. (2008). The phenomena of inner experience. *Conscious. Cogn.* 17, 798–810. doi: 10.1016/j.concog.2007.12.006

Killingsworth, M. A., and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science* 330, 932. doi: 10.1126/science.1192439

Klinger, E. (1999). *Thought Flow: Properties and Mechanisms Underlying Shifts in Content.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Lehmann, D., Grass, P., and Meier, B. (1995). Spontaneous conscious covert cognition states and brain electric spectral states in canonical correlations. *Int. J. Psychophysiol.* 19, 41–52. doi: 10.1016/0167-8760(94)00072-M

Linkenkaer-Hansen, K., Monto, S., Rytsälä, H., Suominen, K., Isometsä, E., and Kähkönen, S. (2005). Breakdown of long-range temporal correlations in theta oscillations in patients with major depressive disorder. *J. Neurosci.* 25, 10131–10137. doi: 10.1523/JNEUROSCI.3244-05.2005

McCrae, R. R., and Costa, P. T. (1994). The stability of personality: observations and evaluations. *Curr. Dir. Psychol. Sci.* 3, 173–175. doi: 10.1111/1467-8721.ep10770693

McVay, J. C., and Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychol. Bull.* 136, 188–197. doi: 10.1037/a0018298

Moffic, H. S., and Pyakel, E. S. (1975). Depression in medical in-patients. *Br. J. Psychiatry* 126, 346–353. doi: 10.1192/bjp.126.4.346

Montez, T., Poil, S. S., Jones, B. F., Manshanden, I., Verbunt, J. P., van Dijk, B. W., et al. (2009). Altered temporal correlations in parietal alpha and prefrontal theta oscillations in early-stage Alzheimer disease. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1614–1619. doi: 10.1073/pnas.0811699106

Raichle, M. E. (2011). The restless brain. *Brain Connect.* 1, 3–12. doi: 10.1089/brain.2011.0019

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.1177/1359105310383604

Sarnthein, J., Stern, J., and Aufenberg, C. (2006). Increased EEG power and slowed dominant frequency in patients with neurogenic pain. *Brain* 129, 55–64. doi: 10.1093/brain/awh631

Schooler, J. W., and Schreiber, C. A. (2004). Experience, meta-consciousness, and the paradox of introspection. *J. Conscious. Stud.* 11, 17–39.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., and Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends Cogn. Sci.* 15, 319–326. doi: 10.1016/j.tics.2011.05.006

Singer, J. L. (1974). Daydreaming and the Stream of Thought: daydreams have usually been associated with idleness and inattentiveness. Now, however, through an empirical research program, their general function and adaptive possibilities are being elucidated. *Am. Sci.* 62, 417–425.

Smallwood, J., and Schooler, J. W. (2006). The restless mind. *Psychol. Bull.* 132:946. doi: 10.1037/0033-2909.132.6.946

Steyer, R., Schmitt, M., and Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *Eur. J. Pers.* 13, 389–408. doi: 10.1002/(SICI)1099-0984(199909/10)13:5%3C389::AID-PER361%3E3.3.CO;2-1

Stoffers, D., Bosboom, J. L., Deijen, J. B., Wolters, E. C., Berendse, H. W., and Stam, C. J. (2007). Slowing of oscillatory brain activity is a stable characteristic of Parkinson's disease without dementia. *Brain* 130, 1847–1860. doi: 10.1093/brain/awm034

Watson, D. L., and Tharp, R. G. (2014). *Self Directed Behavior: Self Modification for Personal Adjustment, 10th Edn.* Belmont, CA: Wadsworth.

Zelenski, J. M., and Larsen, R. J. (2000). The distribution of basic emotions in everyday life: a state and trait perspective from experience sampling data. *J. Res. Pers.* 34, 178–197. doi: 10.1006/jrpe.1999.2275

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in PSYCHOLOGY

The activity view of inner speech

Fernando Martínez-Manrique[1]* and Agustín Vicente[2,3]

[1] Departamento de Filosofía I, University of Granada, Granada, Spain
[2] Ikerbasque: Basque Foundation for Science, Bilbao, Spain
[3] Department of Linguistics and Basque Studies, University of the Basque Country, Vitoria, Spain

We distinguish two general approaches to inner speech (IS)—the "format" and the "activity" views—and defend the activity view. The format view grounds the utility of IS on features of the representational format of language, and is related to the thesis that the proper function of IS is to make conscious thinking possible. IS appears typically as a product constituted by representations of phonological features. The view also has implications for the idea that passivity phenomena in cognition may be misattributed IS. The activity view sees IS as a speaking activity that does not have a proper function in cognition. It simply inherits the array of functions of outer speech. We argue that it is methodologically advisable to start from this variety of uses, which suggests commonalities between internal and external activities. The format view has several problems; it has to deny "unsymbolized thinking"; it cannot easily explain how IS makes thoughts available to consciousness, and it cannot explain those uses of IS where its format features apparently play no role. The activity view not only lacks these problems but also has explanatory advantages: construing IS as an activity allows it to be integrally constituted by its content; the view is able to construe unsymbolized thinking as part of a continuum of phenomena that exploit the same mechanisms, and it offers a simple explanation for the variety of uses of IS.

Keywords: inner speech, format view, activity view, consciousness, unsymbolized thinking, phonological representation, action prediction

INTRODUCTION

Inner speech (IS) is typically characterized as the experience of silently talking to oneself. It is reported as phenomenologically different from other experiences such as visual images, emotions, or the controversial phenomenon of unsymbolized thought (Hurlburt and Akhter, 2008). In this paper we distinguish two general approaches to IS—what we will call the "format" and the "activity" views. These approaches hold different theses about what elements are more relevant to characterize the phenomenon. As we will see, the format view regards IS chiefly as a certain product with certain format features, whereas the activity view emphasizes its properties as an activity. These may appear as mere differences in emphasis—after all, the format view may readily accept that IS is an activity and the activity view does not deny that there is a format involved. Yet the reason for their respective emphases lies in the fact that they have distinct commitments to what is central of the phenomenon. In particular, we will see that the two approaches have different views concerning the cognitive functions of IS, especially whether IS is or is not necessary for conscious thinking.

These are, in general, philosophical approaches, yet empirically well-informed ones. We are aware that, on the one hand, as a verbal phenomenon, a good account of IS will ultimately depend on precise models of linguistic production and comprehension; and that, on the other hand, as a cognitive phenomenon, a plausible account of IS requires more data than we presently have.

However, it is useful to bring to the light the commitments and consequences of holding a certain general view of what IS actually is. In particular, it helps for the methodological assessment of what aspects of the phenomenon it is worthwhile to investigate. In this paper we spell out the differences between the format and activity views, and defend the advantages of the latter.

THE FORMAT VIEW OF INNER SPEECH

The format view is attributable to most authors who have written about the functions of IS in the last two decades[1]. In its strongest form,[2] it can be characterized by the following three theses:

(i) the strong consciousness thesis: IS is necessary for conscious thinking;
(ii) the format thesis: in IS we recruit a representational system because of its features as a format;
(iii) the product thesis: IS consists in some output of the linguistic production system, typically strings of phonological representations.

The first thesis is about the *role* of IS. If "thinking" is roughly understood as any cognitive event that involves the manipulation

---

[1]Exceptions are Vygotskyans like Fernyhough (2009) and Hurlburt et al. (2013).
[2]Along the paper we will introduce a number of weaker versions of the view, which relax one or more theses so as to answer a particular challenge.

or tokening of propositional contents, the thesis says that doing any of this consciously requires the presence of IS. The second thesis is about the *nature* of IS. It says that what is essential for something to count as IS is that it is formatted in a certain way. The third thesis provides a further specification of the kinds of representations involved in IS.

The first and second theses are two sides of the same coin: it is claimed that in IS we recruit a format with certain features because those features open the possibility to have conscious thoughts[3] at all. Different authors have focused on different features, such as digitality, or context-independence (Clark, 1998), perceptuality and introspectability (Jackendoff, 1996, 2012; Prinz, 2011, 2012; Bermúdez, 2003), and predicative structure (Bermúdez, 2003). To take one example: Jackendoff, and Prinz following him, holds that "pure" conscious thinking is impossible for architectural reasons: we can be conscious of intermediate level representations (like 2.5D representations in the visual system), but never of basic-level or higher-level representations, such as concepts or spatial 3D representations. Thus, if we want to have conscious thoughts, we have to use a representational format that has the right kind of representations. Images are good, but phonological representations are much better, given that phonological representations can vehicle many more kinds of thoughts (about the future or past, about *abstracta* and *possibilia*, about relations, etc.).

These considerations lead Jackendoff to the product thesis, i.e., that IS is constituted by strings of phonological representations or structures[4]. One may wonder, however, how central the product thesis is for the format view, and how specific its commitment to a certain type of product is. With respect to centrality, one may contend that the view does not need to regard IS as constituted solely by phonological representations[5]. Surely IS appears as a content-carrying format so it is also constituted by a semantic component. Moreover, the general approach can also be formulated in a way that is compatible with the idea that IS is an action: the action of producing strings of inner linguistic items (mainly) with the purpose of bringing our thoughts to consciousness. In fact, sometimes Carruthers (2011) comes close to presenting IS in this way, so depicting him as endorsing the format view can seem contentious. The difference between this view and what we will call the activity view would perhaps appear as a matter of emphasis and degree.

However, Carruthers (2014), as Jackendoff, Prinz, or Bermúdez, does put the focus on the product and its properties[6]. It has to be noted, on the other hand, that many authors who are not particularly concerned with the issue of the role of IS in conscious thinking, also take IS to be a product (Pickering and Garrod, 2013). That is, it seems to be customary to think of IS as a product and not as an activity of some kind. With respect to the commitment to a specific kind of product, one may observe that there are different kinds of phonological representations. We can distinguish at least articulatory, phonemic, and acoustic phonological representations. We may think that the activity of inner speaking makes use of all three kinds of representations. However, does IS consist in all of them? If IS is characterized in product terms, it seems that IS has to be strings of phonological *acoustic* representations. There are two reasons to support this claim. Firstly, if the format has to be introspectable/perceptual, it seems that only acoustic representations can do the trick, given that neither articulatory nor phonemic representations are introspectable according to his account (see above). So, following Jackendoff, Prinz states that speech sounds, where he includes silent speech, "are experienced at a level that lies above the buzzing confusion of unfiltered sound waves but below the level of phoneme categories" (Prinz, 2012, p. 69).

Secondly, some authors believe that IS as a product makes thoughts conscious because IS is a prediction issued on the basis of an afterward aborted motor action (see Carruthers, 2011; Pickering and Garrod, 2013). Subjects give instructions to produce a certain linguistic item; these instructions are converted into motor commands; and then the command is aborted, but not before an efference copy is sent to the forward models, which issue a prediction about the sensory incoming signal corresponding to the aborted motor command. If this is what IS ultimately consists in, i.e., the prediction of an incoming sensory signal, then, arguably, an instance of IS has to be an acoustic representation, since the prediction represents sounds (not phonemes or articulations).

Be it as it may, we are ready to accept that the association between the strong consciousness and the format theses is more central for the format view than the product thesis, and that any commitments to a certain kind of product typically arise as a consequence of endorsing the two former theses. Indeed, it is only by relaxing these theses that a defender of the format view will be able to deal with some of the challenges for that view that we are going to present.

PROBLEMS FOR THE FORMAT VIEW

We want to present three general problems we see related to the format view—general in the sense that they stem from endorsing its theses (i) and (ii) (strong consciousness and format). First, it has to deny the phenomenon of "unsymbolized thinking" (UT; Hurlburt and Akhter, 2008). Second, it cannot easily explain how

---

[3]As the notion of thought has different uses in the literature, let us spell out the properties that matter for this paper:

(i) A thought is a mental state with propositional content.
(ii) It can be individuated from other thoughts in terms of its content.
(iii) It can be unconscious or conscious, so it is possible to have the same thought in both modalities.

A conscious thought, thus, is a conscious mental state with propositional content, for instance, a conscious judgment that *p*.

Finally, even if "having a thought" and "thinking a thought" could point toward passive/active occurrences of thought, this is a distinction that we do not discuss in this paper so we will use both expressions interchangeably.

[4]See, e.g.: "[Chomsky] has fallen into the trap (...) of believing that inner speech is thought, rather than (as I will argue) the phonological structure corresponding to thought" (Jackendoff, 2007, p. 70), and "conscious thought gets its form (...) from the inner voice, the verbal images of pronunciation" (Jackendoff, 2012, p. 103).

[5]We owe this objection to a referee.

---

[6]"Especially important (...) are the auditory images that result from off-line activation of instructions for producing speech, which result in auditory representations of the speech act that would normally result, in so-called 'inner speech'" (Carruthers, 2014, p. 149).

IS makes thought-contents available to consciousness (Jorba and Vicente, 2014). Third, it may have problems in accounting for the variability of uses of IS. In addition to these general problems, we will finally examine a particular rendering of the IS-as-a-product idea, namely, the suggestion that IS is an acoustic representation that predicts an incoming sensory signal—a suggestion that has some problems of its own.

THE PUZZLE OF UNSYMBOLIZED THINKING

Using the method of Descriptive Experience Sampling, Heavey and Hurlburt (2008) reported that people claimed to experience inner episodes in which they had the feeling of "thinking a particular, definite thought without the awareness of that thought's being conveyed in words, images, or any other symbols" (p. 802). For instance, someone could report her experience as wondering whether a friend would be driving his car or his truck but with no words carrying this specific content, and no images of the friend, the car or the truck (Hurlburt and Akhter, 2008, p. 1364). According to their results this kind of "unsymbolized thinking" occupies around an average of 22% of our conscious life (Hurlburt and Akhter, 2008; Hurlburt et al., 2013).

Unsymbolized thinking is not an uncontroversial phenomenon. Even though there are other strands of research that point toward a distinctive phenomenology for propositional thought (Siewert, 1998; Pitt, 2004), its characterization is elusive. For instance, Hurlburt and Akhter (2008) portray it mostly in a negative way, holding that "unsymbolized thinking is experienced to be a *thinking*, not a feeling, not an intention, not an intimation, not a kinesthetic event, not a bodily event" (p. 1366). In this paper we do not wish to enter the debate concerning the evidence for UT. Rather, the point we want to make is conditional: *if* UT is a genuine phenomenon to explain, it poses a serious problem for the format view. This view claims that we recruit IS so that we can have conscious thoughts—otherwise, we would not be able to think consciously. But if it is possible to have conscious thoughts without the presence of IS then the format view's claim is simply false. Indeed, its best strategy is simply to deny this phenomenon. In this vein, Carruthers (2009) argues that UT may be a result of confabulation: people report thinking without words or images, but they may be actually using words and/or images, or they may not be really thinking (e.g., they think that they were thinking about what product to buy, but in fact they were only looking at the different products). Hurlburt et al. (2013), in contrast, suggest that confabulation probably goes the other way around: we engage in more UT than that 22% average, but as we tend to identify thinking with innerly speaking, we tend to report using words when in fact we are not using them.

To repeat, any view that endorses both the strong consciousness and the format theses will hold that, in fact, IS is the form that conscious propositional thinking adopts[7], so inasmuch as UT is propositional it is simply impossible. However, it is possible to construe weaker versions of the format view in which UT appears as a more tractable phenomenon. In particular, one may drop the

strong consciousness thesis and hold that IS is not *necessary* to have conscious thoughts. IS would be only a good way, possibly the best, to make thoughts conscious, but there are other ways to do so. Perceptual theories of consciousness (Prinz, 2011) are a good candidate for this weaker version. These theories claim that a thought always needs a certain perceptual format in order to be conscious, and that "even high-level perceptual states and motor commands are inaccessible to consciousness" (Prinz, 2011, p. 174). IS constitutes a variety of such a perceptual format but there could be others. In particular, there could be *non-symbolic* perceptual vehicles, like emotions, or bodily feelings. Following this path, there is a chance to account for UT without denying the phenomenon: an unsymbolized thought would be a thought that is cashed-out in some non-symbolic perceptual format.

There are problems for such an account. A first problem is that it is not clear that it actually fits the characterization of the phenomenon offered by researchers of the phenomenon. Recall that Hurlburt and Akhter (2008) reject that UT is experienced as a feeling, intention, intimation, kinesthetic or bodily event, adding that people "confidently discriminate between experiences that are thoughts (…) and experiences that are feelings (…) or sensory awareness" (p. 1366). This seems to leave very little room to manoeuver for a perceptual account of UT. Now, one may protest that Hurlburt and Akhter's (2008) positive characterization of the phenomenon is somewhat wanting and that there is perhaps a different kind of perception behind it. So let us focus on a second problem that seems to be more pressing for the perceptual account, namely, the problem of accounting for the specific semantic content of the unsymbolized thoughts that subjects report.

If UT is a genuine phenomenon the only positive characterization we have is that subjects claim to be experiencing definite thoughts[8]. So any account of the phenomenon will have to respect this characterization. Consider the unsymbolized wondering whether a friend would be driving his car or his truck. What sort of perceptual experiences could carry that content? If the subject were engaged in an experience of IS the answer would be straightforward: it is the content of a mental sentence. But non-symbolic perceptual experiences, such as certain feelings associated to your friend and his truck, appear as unsuitable for that task. Certainly, in Prinz's view (e.g., in his theory of emotions, Prinz, 2004) feelings can have intentional contents, but they do not seem to be so nuanced to include the specific content of a thought such as the subject's wondering. Prinz's suggestion of treating propositional attitudes in terms alike to emotions (Prinz, 2011) can help with respect to the "attitude" part, i.e., it might be the case that what distinguishes "wondering whether

---

[7]See, e.g., Bermúdez (2003, pp. 159–160): "[A]ll the *propositional* thoughts that we consciously introspect (…) take the form of sentences in a public language" (his emphasis).

[8]A referee points out that Vygotsky's distinction between a natural and a cultural line of development is relevant to the question of UT. Those two pathways to thought could result in kinds of thinking with different properties, and UT could occur in both of them, so its analysis would have to take into account the distinction. We agree that this might be the case and insist that a definite characterization of UT is still lacking. In this paper we will limit ourselves to the minimal characterization offered by Hurlburt et al. (2013)—i.e., UT as thought with propositional content and a "proprietary" phenomenological basis—and we sketch a proposal that would link it to the cultural line—see Section "The Relation Between Inner Speech and Unsymbolized Thinking."

p" from "doubting that *p*" is a certain emotion-like feeling that accompanies the thought. Yet this feeling does *not* account for the experience of the content *p*, so something else must back the latter experience. Given the problems of attaching specific propositional contents to visual or other non-verbal sensory elements (more on this in the next section), Prinz does not seem to have other resources than imaged sentences. Therefore UT appears for him as unlikely as for other defenders of the format view.

Perhaps a way out of this problem is to claim that the non-symbolic perceptual format is recruited, but not to broadcast thought-contents, but to *prompt* them. That is, perceptual experiences would not be used as *vehicles* of the content but only as means to focus our attention or to keep track of our thought processes. Conscious thinking may thus be unsymbolic in Hurlburt's sense, even though many times unsymbolic conscious thinking uses perceptual scaffolding. Yet, this alternative view seems full of problems.

The format view provides an account of how IS is generated, and tries to explain how IS makes conscious thinking possible. Yet it has no explanation about conscious thinking which is not supported by IS—the prompting model appears as an *ad hoc* addition to it. If we take Carruthers's model as a paradigm of the format view (see below), it is clear that the model is not made to explain that IS *prompts* conscious thoughts, but to explain that IS *vehicles* conscious thoughts. Producing a string of phonological representations with contents attached *is* having a thought, according to the model, whereas the prompting model would say that producing a perceptual surrogate—verbal or otherwise—just *facilitates* having a thought in consciousness, the relation between the prompt and the content being arbitrary.

Finally, the format view also seems to account for the sense of agency related to mental phenomena inasmuch as it construes them as motor phenomena. For instance, in Carruthers's model agent awareness is explained on the basis of the production of imagery that engages the forward model system. The details of how the sense of agency emerges are not clear[9], yet it seems that the prompting model cannot explain why prompted thinking would feel as our own thinking. The only thing that one would feel as his own would be the prompt.

HOW THOUGHT-CONTENTS ARE AVAILABLE TO CONSCIOUSNESS

Even if one disputes the evidence for UT, the format view still has the problem of explaining how thought-contents are available to consciousness (see Jorba and Vicente, 2014, for extended discussion). Any account of conscious thinking has to explain how thought-contents become access-conscious[10]. Defenders of the format view hold that by producing strings of phonological

representations we bring thought-contents to consciousness. Yet, it is not explained how this is done. It seems that by speaking to ourselves we become conscious of the phonological structure of our IS. How does this kind of consciousness explain consciousness of meanings, or contents? Remember that, on some accounts, like Jackendoff's, conceptual structures and therefore meanings and propositional contents, are necessarily unconscious. The question then is: how do these structures or representations become conscious, at least, access-conscious, by virtue of making phonological structures conscious?

Clark (1998), as well as Bermúdez (2003) and Jackendoff (1996, 2012) propose that phonological representations convert propositional contents into objects that become present to the mind's eye. However, it seems that converting a propositional content into an object one can "look at" only enables subjects to know what they are thinking, not to think those thoughts consciously. Instead of making them aware of a certain propositional content *p*, and so to consciously believe or judge that *p*, this mechanism makes them aware that *they are thinking* that propositional content, i.e., that they are believing or judging that *p*. Objectifying seems to give the subject metarepresentation, but not ground-level conscious thinking.

Let us clarify this point in terms of Clark's position. Clark (1998) presents his view as a development of Vygotsky's ideas about IS Vygotsky (1987). However, the role he envisions for IS is very different from Vygotsky's emphasis on the role of IS in self-regulation and executive on-line control, as well as in planning more or less immediate actions—that is, not planning a summer trip, but planning how to solve the Tower of Hanoi task. Vygotskyans typically hold that IS helps us focus our attention on what we are doing, whereas Clark et al. hold that it makes possible for us to focus on what we are thinking. Vygotskyans point out that IS is involved in, inter alia, executing an action step by step. This means that IS enables us to do whatever we are doing in a conscious mode. We monitor our behavior by consciously thinking "this goes here," "this goes there," "if this goes here, then that goes there," etc. In contrast, Clark's model is a model not of behavior control or monitoring, but apparently of metacognition, i.e., of knowing what we think. We believe there is a difference between saying that IS helps us to have conscious thoughts, which are used to monitor and control our behavior, and holding that IS makes us aware of what we are thinking, so that we are able to think about our thinking.

Perhaps Clark, Jackendoff and Bermúdez do not intend their account to have the narrow scope we are ascribing to it[11]. However, the model they propose seems to only be able to explain

---

[9]As we will see in Section "The Relation Between Inner Speech and Unsymbolized Thinking," the view that IS is an incoming sensory signal seems to fare better in this respect, for it involves comparisons, which many regard essential to the generation of self-attribution (see Frith, 2012).

[10]As it is well known, the distinction between phenomenal and access consciousness was first introduced by Block (1995). Phenomenal consciousness is defined in terms of what-it-is-likeness or experience, and Access consciousness is characterized as information being available to the direct rational control of thought and action.

[11]However, see Clark (1998, p. 171): "[P]ublic language (. . .) is responsible for a complex of rather distinctive features of human thought viz, our ability to display *second order cognitive dynamics*. By second order cognitive dynamics I mean a cluster of powerful capacities involving self-evaluation, self-criticism and finely honed remedial responses (. . .) This thinking about thinking, is a good candidate for a distinctively human capacity (. . .) Jackendoff (. . .) suggests that the mental rehearsal of sentences may be the primary means by which our own thoughts are able to become objects of further attention and reflection." See also Bermúdez (2003, p. 163): "We think about thoughts through thinking about the sentences through which those thoughts might be expressed."

how IS gives us knowledge of what and how we think. Let's say that by using sentences of our language, we are able to have some kind of object before our minds. What do we gain with that? Presumably, we only gain knowledge about what we are thinking. We "see" the sentence, get its meaning, and reach the conclusion "ok, I'm thinking that *p*." This knowledge about what and how we are thinking may be very useful, of course, but we would say that this is only a use of IS, among many others[12]. The account, in any case, does not explain how thought-contents are made access-conscious.

In this respect, Carruther's (2011, 2014) idea that thought-contents are bound into strings of phonological representations and broadcast along with them fares much better. For according to this idea, thought-contents as such make it into access-consciousness by being bound to formats which are both phenomenal and access-conscious: "there is every reason to think that conceptual information that is activated by interactions between mid-level areas and the association areas (. . .) gets bound into the content of attended perceptual states and is broadcast along with the latter. Hence we don't just see a spherical object moving along a surface, but a tomato rolling toward the edge of the counter top; and we don't just hear a sequence of phonemes when someone speaks, but we hear what they are saying; and so on" (Carruthers, 2014, p. 148).

What is not clear in this view is how the binding process takes place, especially given that, according to Carruthers, what we do in order to extract the meaning of an IS episode is to interpret an already conscious phonological representation by means of the usual comprehension mechanisms. According to Langland-Hassan (2014), however, the only content that can be bound into an episode of IS is of the kind: the semantic meaning of this episode of IS is such and so. That is, the content bound into the string of IS would not be about the world, as it should be, but about the very string[13]. The reason is, basically, that phonological representations represent acoustic properties, while semantic representations represent the world. Langland-Hassan argues that there is no way to fit these different kinds of representations into a single item.

There are perhaps reasons to resist this idea. If one regards representational content as the information that a representation conveys, it is clear that a representational instance can convey different kinds of information. A phonological representation may represent sounds but it is by means of this acoustic information that it also represents certain semantic information. That is, in a nutshell, Prinz's position (Prinz, 2011, 2012). Prinz argues that consciousness requires attention to sensory representations. These representations are "images generated from stored concepts [that] inherit semantic properties from those concepts" (Prinz,

2011, p. 182). IS constitutes a particularly important kind of images, i.e., linguistic images, which carry information *both* about acoustic properties and semantic content. In this respect, Prinz's theory seems to eschew Langland-Hassan's criticism: causal-informational chains are responsible of keeping the different sorts of information attached to the same sensory representation, so the binding problem may not arise.

However, Langland-Hassan's analysis also raises another concern: those different contents have different functional or inferential roles to play. Acoustic information will play a role in inferences having to do with the representation's sound, while semantic information will be routinely exploited for reasoning processes having to do with what those words mean. Those inferential roles cannot simply be mixed together. Again, Prinz's view may have a way out of this difficulty: those contents are not attended at the same time. To have conscious thoughts, a subject must have a certain sensory representation in mind *and* attend to it, but nothing precludes that at some times she attends to its sensory properties, and others to its semantic content. So thoughts are available to consciousness simply by attending to the sensory elements related to the semantic representation proper.

We think there is a problem in this position. Compare the case in which a subject is attending to the representation's sensory information with the case in which she is attending to its semantic information. What is the phenomenological difference between both cases in the subject's mind? According to Prinz's perceptual consciousness account, there must be some sensory difference between them, e.g., an accompanying sensory representation. So if the subject is thinking about the representation's acoustic information some acoustic-related representation will be present; if she is thinking about its semantic information, some semantic-related representation will be present.

This account paves the way to an infinite regress. Notice that accompanying representations have to be sensory representations themselves, and the same sort of question can be raised with respect to them: does the subject attend to its sensory or to its semantic information? To distinguish between both cases one must appeal to further distinct accompanying representations, which are sensory representations themselves and which raise the same kind of issue. To put the problem in different terms: if you have a theory in which for a thought to be conscious it must be cashed out in a certain format, then you introduce a gap between the thought's content and the content of the format itself. What makes the thought conscious cannot be simply the format because there is always the question of how that particular format makes that particular thought conscious.

THE VARIOUS FUNCTIONS OF INNER SPEECH

The final problem for the format view we want to mention is that it is not clear how it can account for the variability of uses and of kinds of IS. We use IS in most of the kind of situations where we may use outer, or overt, speech (OS). For instance, IS is used for motivating, encouraging, entertaining, expressing the speaker's emotions or feels, guiding behavior, etc. The main difference is simply that OS can be addressed to someone else whereas IS has to be addressed to oneself. So among the functions of OS that we probably would not find in

---

[12]On the other hand, second-order dynamics and metacognition are probably different phenomena. We can know what we are thinking just by having conscious thoughts: once you think a thought consciously, you also know that you are having that thought. In this respect, thinking is similar to perceiving: when you have a conscious perceptual experience, you thereby also know that you are having that experience. What objectification gives us, we would say, is the ability to reflect about our thinking and to gain control over our higher-level cognitive processes.

[13]In philosophical jargon, the content would be token-reflexive.

normal IS we can count those actions that conceptually require somebody else, like promising and threatening, perhaps—yet IS can include comparable functions, such as warnings. At any rate, this is just a reflection of how the things one can do with language depend on the audience one is addressing but this reveals no important, or deep, functional difference between outer and IS.

When it comes to explaining the plurality of functions of IS, the format view may have a problem. The format view is not committed to claiming that we only use IS for having conscious thoughts. However, apparently, it does propose a story about why IS is recruited and thus seems to commit to a certain idea about the *proper function* of IS: the proper function of IS would be to make conscious thinking possible, while uses of IS not related to conscious thinking would be derivative. Yet it is difficult to see how such derivation would proceed. For instance, if one considers the case of OS, one cannot find an analogous fundamental function. One might appeal to the notion of "communication," arguing that it is akin to the very general function of "focusing someone else's attention on something," or "making someone conscious of something." Yet this is at most a loose way of speaking.

Let us flesh out a general motivation that lends support to the thesis that IS may have a proper, constitutive, function. There is this old conundrum about why someone ought to talk to herself, when she knows in advance what she is going to say. In other words, if one thinks that the semantic content is "already there" before the words are actually uttered, one should not bother to put it in words for oneself. In other words, IS cannot have a communicative function because communication presupposes an informational mismatch between the speaker and the listener, and this mismatch does not exist when both roles concur in the same person. Second, it is not clear that some uses of IS count as communication. For instance, it does not seem to be necessary to characterize self-motivation, or even self-evaluation or self-awareness (Morin, 2011), in terms of communication. It is weird to say that when you motivate yourself with words you are engaging in some act of communication with yourself. If IS does not have a communicative function it must have a function of its own. Which one? A promising response seems to be that IS has a function related to conscious thinking.

Even though this is an alluring motivation, we think it has a basic flaw: it seems to assume that the function of outer speech is *merely* communicative. However, this is not the case. OS can play the same cognitive roles as IS, including the alleged roles related to consciousness. When the mother, helping her daughter to solve a jigsaw puzzle, tells her "this here. . . that there," etc., she is directing her attention to the items and the places, i.e., she is regulating her behavior by talking, just as we are supposed to be doing when we use IS. In principle, anything that we tell in IS could be told in OS, and for exactly the same purposes. So if IS had the function of making thought-contents conscious, it would certainly not be its proper function but a function of speech in general (e.g., in the case considered, we can say that the mother is making her daughter conscious of where the different pieces go, so that the daughter consciously judges that this piece goes here, etc., thus gaining control over the resolution of the puzzle). IS would

not have the communicative function of OS but IS's functions could still be considered as a subset of OS's.

However, this "proper function" commitment may be not essential to the view. It is relatively easy to read authors as endorsing claims about the proper functions of IS—many statements take the form of "we use IS for x," where x is substituted by conscious thinking, system-2 thinking (Frankish, 2010), self-regulation, executive control, or whatever. Yet, it may be non-charitable to read these claims as expressing strong views about proper functions. A more liberal reading is to think that each author has focused on a use of IS and has simply apparently left the rest in the background. We think it is methodologically advisable to start by first detailing the different uses of IS, the different situations where we use it, as well as the different kinds of IS that there may be, but this is a different issue (for examples of this kind of approach, see Morin et al., 2011; Hurlburt et al., 2013). The point now is that defenders of the format view may drop a strong commitment to a proper function of IS and accept a plurality of uses.

However, even if the "proper function" commitment is abandoned, we think that when it comes to account for the uses of IS the format view typically has the order of explanation backward. The story assumes that IS couches thoughts in a certain format, and that, by doing so, those thoughts can be put to new, different uses. Yet the functional order is just the opposite: thoughts are formed and recruited to be put to different uses and, in doing so, they can appear in a certain format. Consider the example of an athlete telling herself motivating words (Hatzigeorgiadis et al., 2011). The athlete does not first form the mental sentence "you can do it" and then use this sentence to motivate herself. Rather, the athlete is engaged in the activity of motivating herself and, in doing so, her motivating thoughts can reach the point in which she hears herself telling encouraging words silently (or even aloud sometimes). Or consider the case of someone deciding to put more money in the parking meter and telling himself "One more quarter? Mmm. . . Can be back in one hour. Better a coffee." The subject is making a decision by means of certain conceptual activity. Some of the elements of this activity—typically the most salient and relevant ones—can emerge to consciousness under verbal control, where they can be put to further uses and lead to new cycles of mental activity. These two examples are cases in which the linguistic production system may be recruited spontaneously so that, so to speak, "words come to our mind" but, of course, we can also *bring* words to our minds by engaging explicitly in linguistic activity. The student preparing a talk may revise innerly some of the sentences she intends to utter, so as to change a few words, decide where to put the emphasis, and the like. Again, the way of describing this is not that she is putting her thoughts in verbal format and then examining them. Rather, she is already engaged in the activity of examining her own thoughts on the matter she wishes to talk about and uses her verbal systems so as to do this in a more precise manner.

On the other hand, endorsement of the format view involves that, even if one abandons the idea of a proper function, one still holds the claim that recruiting a format plays a necessary role in the plurality of functions. Yet some of those functions cast doubts concerning the claim that the format is necessary—let

alone the linguistic format. Think again about IS and motivation, which is amply discussed in the psychology of sports literature (Hatzigeorgiadis et al., 2011). An athlete does not need any kind of particular format to motivate herself: she may tell herself "give it all!!," but she could just as well fix her sight at the finish line and see how close it is, feel how fast her legs are moving, or whatever. She needs perceptual or proprioceptive stimuli, but these do not have to be self-produced (i.e., they do not have to be the result of imagery or IS production).

Finally, the idea that in IS we always recruit a format for a purpose is also open to doubt. There seem to be cases where the only thing we do with IS is add a clearly unnecessary expressive commentary to something that we have done (Hurlburt et al., 2013), like the 'a-ha's, or 'great!'s we tell ourselves after, for instance, having thought hard about something. Would we say that, in these cases, we are recruiting a format with some purpose? Arguably, we would not put it in that way. Moreover, we would probably say that we are using IS with no purpose at all—at least no purpose related to the cognitive activity in question. Yet, non-purposive IS seems to be a problem for the format view however weakly it is construed, for the format view wants that phonological representations are used to perform cognitive functions.

IS INNER SPEECH A PREDICTION?

In this last section about the problems of the format view we want to consider briefly the particular proposal about IS we have mentioned above, namely, that it is a prediction about the linguistic sounds that one would hear if a certain linguistic action had not been aborted. This proposal has some independent appeal, as it construes IS as a species of motor imagery (Carruthers, 2011, 2014). Current theories of motor imagery (Jeannerod, 2006) hold that motor imagery results from aborting the execution of motor commands, and from generating a prediction about sensory and proprioceptive incoming signals. It is appealing, we think, to embed IS in a larger theory about imagery production.

However, the proposal that an episode of IS is a prediction about linguistic sounds does have some problems. One first problem is that it cannot accommodate the intuitive idea that IS is typically experienced as *meaningful*, e.g., when one is engaged in conscious reasoning. This is in contrast with meaning-ignoring instances of IS (e.g., when one repeats some linguistic items mentally so as to memorize them—we will call these cases "meaningless" for short). We would say that when we talk about IS in contexts like the present one, we are only talking about meaningful IS. However, the way the format view prefers to individuate IS does not need semantics, meaning or content—or if it has a role for semantics, it is a secondary one, ancillary to the format's properties. So both meaningful and meaningless instances of a string of phonological representation could count as the same type of IS.

The proposal also seems to have problems to deal with data which apparently show that IS may contain errors which are recognized as such (Oppenheim, 2013), because, prima facie, a prediction issued on the basis of an efference copy is not monitored; rather, its proper function is monitoring production. A related, and complicated problem, is that the proposal excludes the cur-

rently widely accepted idea that passivity phenomena in cognition (auditory verbal hallucinations (AVHs) and thought insertion) may derive from a misattribution of IS (e.g., Ford and Mathalon, 2004; McCarthy-Jones, 2012; see also Langland-Hassan, 2008, for a revised version in terms of a filtering/attenuation deficit)[14]. This latter idea seems to require that IS is an *incoming signal* against which a prediction is compared, rather than this very prediction. That is, misattribution (as error checking) is only possible when there is comparison, which in turn requires a prediction *and* an incoming sensory signal. If the only product we get from inner speaking is a sensory/acoustic prediction, then it is mysterious how we could self- or other-attribute it (see, however Vicente, 2014 for development and criticism of the idea that IS is an incoming sensory signal). It seems that both error checking and misattribution require that IS is *not* a prediction about linguistic sounds issued by the forward models.

THE ACTIVITY VIEW OF INNER SPEECH

The view we want to argue for stresses the *activity* of innerly speaking, instead of the format of IS. This view is not without precedent. For instance, the emphasis on activity is a key ingredient in the Soviet school to which Vygotsky belongs (Kozulin, 1986; Guerrero, 2005) and many contemporary Vygotskyans understand language as activity-based (Carpendale et al., 2009) and IS as an internalization of this activity. Other recent approaches that characterize IS as preserving some feature of linguistic activity—and not merely linguistic format—include Fernyhough (2009), who conceives of language as inherently dialogical, or Hurlburt et al. (2013), who commend the use of inner *speaking* to avoid regarding IS as mere representational product.

In relation to the format view we depict in this paper, our idea of an activity view of IS rejects both the format and the strong consciousness theses associated to the former. With respect to the format thesis, it claims that in IS we do not recruit a format, be it perceptual, predicative, or whatever. At most, we could say that we recruit a linguistic activity, though we think using the notion of recruitment mischaracterizes the view: we do not properly recruit the activity of speaking; we just speak, although innerly. With respect to the consciousness thesis, the view denies that IS is necessary for thinking consciously, or that IS is *for* thinking consciously (i.e., that its proper function is conscious thinking). Rather, the activity view adopts a pluralistic stand: IS has almost as many functions, or uses, as we can discover in OS, none of which should be singled out as its proper function.[15]

If we observe our own IS we will see that, in effect, IS is put to use in many different circumstances: self-expression, motivation,

---

[14]However, Langdon et al. (2009) dispute this claim on the basis of studies with schizophrenic patients. Comparing their AVH and IS, they found no similarities between their phenomenological characteristics—similarities which arguably ought to be present if AVHs derive from IS.

[15]The continuity of function between inner and outer speech is a typical assumption in those that understand IS as inheriting the functional roles of the private speech from which it originates (see reviews in Berk, 1992; Winsler, 2009). Relations between inner and outer speech are also currently the focus of attention of empirical research in terms of parallelisms and differences in the linguistic subsystems responsible for their respective processing—e.g., the comprehension and production systems (Vigliocco and Hartsuiker, 2002; Geva et al., 2011). Those topics exceed the purposes of this paper.

evaluation, attention-focusing, self-entertainment, fixing information in memory, preparing linguistic actions, commenting on what we have done, accompanying our thoughts, etc.[16]. There seems to be no deep difference between reasons why we talk to ourselves and reasons why we talk to someone else: we talk to express ourselves, to motivate others, to evaluate events or subjects, to help people to find places, to regulate their behavior, etc. Moreover, there seems to be no deep difference between the way we talk to ourselves and the way we talk to someone else. For instance, if we want to motivate our favorite athlete, we may tell her "come on!," "you're the best!," that is, the kinds of things she may be telling herself. If we want to help someone to get to a certain destination, we may use a map and tell him "you go here, then there. Go straight this way, turn here," etc. That is, we insert linguistic fragments within the background provided by the map, which is what we do when we mix mental maps and IS in orientation.

There are also parallels between the cases in which IS and OS appear in longer, more elaborate linguistic constructions vs. those in which they appear condensed or fragmentary. For instance, when we talk about ourselves, or about a certain person or event that concerns us, we typically use full sentences, and elaborate a narrative, just as we do when we get introspective about ourselves, other people, or certain events. On the other hand, our speech appears as condensed or fragmentary if we are regulating someone else's behavior on-line: the adult that helps his kid to complete a jigsaw puzzle, tells him "this piece here. Square there? Sure? Where is a triangle missing? No. Yes," etc. As has been long highlighted by Vygotskyans, IS, when put to this kind of use, is equally typically condensed[17]. This suggests that using IS is, basically, innerly *speaking* (see also Hurlburt et al., 2013).

The activity view we propose is in clear contrast with the strongest versions of the format view, i.e., those which hold that IS is for conscious thinking, and that IS is necessary for conscious thinking because we need a certain format to get thought consciousness. However, in the discussion of the format view we have considered weaker versions of it. A weak version of the format view, for instance, could simply claim that we produce phonological representations to better do a variety of things, from conscious thinking to motivation. The activity view and this weak version of the format view do not look that different in principle.

However, there are reasons to prefer to categorize IS as an activity *tout court* rather than in terms of a format. First, labeling IS as an activity fits better the natural description of IS as speaking, and not as producing phonological representations (even if phonological representations are produced). Second, the notion of activity underscores the functional continuity between outer and IS in a more natural way than the format view. As we explained, the format view typically begins by focusing on a function that is putatively exclusive to IS, i.e., thought consciousness.

The consequence is that it tells apart outer and IS—the former is an instrument of communication, the latter of cognition. Even if one relaxes the account to make it sensitive to the plurality of uses of IS, it tends to consider these uses as solutions to particular cognitive demands. The activity view, in contrast, regards them as predictable effects of internalizing OS and its different functions.

Be it as it may, the view we want to propose deserves the label "activity view" on further grounds, which mark a stronger contrast with the format approach. We claim that IS, as speech in general, is characterized as a *kind of action*, namely, an action that consists in expressing thoughts. In philosophical parlance, this means that IS is *individuated* in terms of the action it is, i.e., that it is distinguished from other mental phenomena attending to what the person (or the person's mind) is doing. This excludes that IS should be individuated in terms of its product qualities, e.g., its properties as a string of phonological representations.

The question of how to individuate IS is not a mere metaphysical point but has important methodological consequences about how one should approach its study or what sorts of mental mechanisms are relevant for it. For instance, by laying the focus on the action of speaking, it is quite natural to try to understand IS in terms of all the representations that are mobilized in speech, i.e., semantic, syntactic, maybe articulatory, etc. As we argued in Section "How Thought-Contents are Available to Consciousness," in the format view the semantic properties of an instance of IS appear as something that one has to bind to it—not as something inherently constitutive of it—raising concerns about how the binding takes place. In contrast, for the activity view the act of innerly speaking begins with a prior intention to express a certain thought that can get more and more specific, until it reaches the level of motor commands. The representations involved in the activity—from conceptual to phonological—form an integrated system, and the ultimate format's properties have no privileged role in accounting for the phenomenon and its functions.

ADVANTAGES OF THE ACTIVITY VIEW

We hold that the activity view has several advantages over the format view. In this section we will develop a particular proposal about how the activity view can explain certain phenomena. The activity view, as we have presented it, is rather liberal in its commitments. Thus, it is compatible with what we have said so far to hold that we do not have to bind thought-contents to phonological representations: it can be said that we interpret our IS just as we interpret OS, i.e., by means of the linguistic-plus-pragmatic system. It is also compatible with the view to have it that, although we sometimes use IS in certain activities where conscious thought is involved, conscious thinking is possible without IS. That is, the spirit of the activity view is consistent with a general model of conscious thinking which has it that conscious thinking is typically unsymbolized: sometimes we speak to ourselves as an aid—but in that case we cannot be said to be thinking in IS, and sometimes we engage in conscious thinking directly (for a sketch of this view, see Jorba and Vicente, 2014).

Here we will pursue a different view according to which predictions issued on the basis of high level intentions play a prominent role both in binding contents into phonological representations (or in making IS meaningful) and in explaining UT. On the one

---

[16]See Morin et al. (2011) for a study that taps the variety of functions of IS.

[17]Vygotsky (1987) and followers have typically been concerned with the use of IS in self-regulation, as they have been particularly concerned with the moment kids start internalizing not just speech but social life in general. Yet, the on-line regulation of behavior is just one function of speech among many others, and it seems that there is no reason why speech should be used only for that purpose when it gets converted into IS.

hand, we regard this proposal worth exploring because it seems to be able to unify apparently different phenomena. On the other, it is the only proposal that we can think of right now which could explain the nature of UT and the sense of agency attached to it. In all, we think it has more explanatory power than the view we have just mentioned.

INNER SPEECH AS MEANINGFUL

As we said above, there is a distinction between meaningful IS (involved in the panoply of functions we talked about in the previous section) and meaningless IS (which we use, for instance, in order to simply retain uninterpreted items). If one regards IS as the strings of phonological representations generated by linguistic productions systems, the consequence is that IS is not meaningful *per se*. In other words, the distinction between meaningful and meaningless instances of IS has to be accounted for in some additional mechanism, for instance, an attentional mechanism that puts the focus either on the semantic or the phonetic information of the representation—which, as we argued, poses an explanatory problem. In contrast, the activity view regards meaningful and meaningless IS as different kinds of actions. It is not the case that a subject produces a certain phonological representation and then puts it to different uses, or under different attentional processes. Rather, the very production of the phonological representation starts with different intentions that mobilize different sets of representations, e.g., in the case of meaningless IS semantic representations are simply not mobilized to begin with. In concordance with this approach, we think that the notion of inner *speech* proper corresponds only to its meaningful instances[18].

Another related advantage is that, by insisting on the idea that IS is inherently meaningful, the activity view easily avoids one aspect of the binding problem we mentioned in Section "How Thought-Contents are Available to Consciousness." As we pointed out above, it is not easy to see how something that represents sounds may also (semantically) represent the world. So if we individuate IS in terms of format properties, we have to explain how content gets bound to it. In contrast, according to the view we are proposing, IS proper is meaningful, and content is an integral part of IS episodes—it does not appear as something "external" that one somehow attaches to represented sounds. Moreover, we are in a position to claim that the content of an IS episode is not the content that phonological representations could eventually encode, but the content that the subject intends to express. In other words, the activity view agrees that in IS the content eventually adopts a certain format, but the specific properties of the format are secondary to explain the phenomenon.

This issue turns out to be particularly important when we consider condensed or fragmentary IS: a linguistic fragment (say, "the ball!") can be used to express many different thoughts (that I lost the ball, that you lost the ball, that we left the ball at home...). Most utterances, if not all, can express different thoughts, depending on the circumstances, but fragments are especially ambiguous (Vicente and Martínez-Manrique, 2005,

2008; Martínez-Manrique and Vicente, 2010). Now, how can we say that the string of phonological representations that constitute "the ball!" means, e.g., that we left the ball at home? It only conveys this specific content if we take into account not the representations themselves but the intentions of the speaker. It seems to us that this sort of response is not so easily available for format views. In particular, the position we attributed to Prinz above may have trouble in explaining how the intended content (i.e., the content subjects want their words to have in a particular occasion) gets bound into the phonological output.

BINDING AND THOUGHT CONSCIOUSNESS

There is another aspect to the binding question, however. In fact, it is this other aspect that occupies Carruthers (see How Thought-Contents are Available to Consciousness). Recall that Carruthers resorts to binding in order to explain how thought-contents become access conscious. His view is that thought-contents can be bound into phonological representations and be broadcast together with them. Carruthers, thus, is not so much concerned with how phonological representations have meaning as with how this meaning is broadcast and made available to higher-level cognition. That is, Carruthers's binding account is a response to this latter issue. The question, then, is: can the activity view do better than Carruthers's version of the format view in this respect? We want to argue that it can.

In motor imagery, as well as in motor acts, the brain issues efference copies and predictions, which are used to monitor and eventually correct actions on-line, as well as to confirm authorship (Jeannerod, 2006). It is not yet clear how the sense of agency arises (see The Puzzle of Unsymbolized Thinking), but it seems likely that it is linked to the good functioning of the forward-models system of efference copies and predictions. Now, less is known not only about so-called mental actions, but also about how the system handles higher-level intentions. However, one can claim that the system does not only receive efference copies from motor commands and issue predictions about incoming sensory signals; it also has to receive efference copies from higher-order intentions and to make predictions on that basis (see Pacherie, 2008).

The architecture for the comparator system proposed by Pacherie (2008) involves a hierarchy of intentions and predictions. This allows her not only to explain how it is possible to monitor the execution of higher-level intentions, but also to provide an account of the different components of the sense of authorship. Pacherie distinguishes three levels of intentions: distal, proximal, and motor intentions (motor commands). Distal intentions are about the goal of the action; proximal intentions are about the here-and-now execution of the distal intention; and motor intentions are about the movements of the body that will eventually realize the proximal intention. As she says, each kind of intention deals with a particular type of representation: "The contents represented at the level of D-intentions as well as the format in which these contents are represented and the computational processes that operate on them are obviously rather different from the contents, representational formats and computational processes operating at the level of M-intentions" (Pacherie, 2008, p. 192). According to her, distal (D) intentions work with propositional/

---

[18]We are aware that one can find a variety of uses for the label "inner speech" in the literature, and we do not mean to legislate the usage of the term. We just want to lay the emphasis on the distinct sort of phenomena that meaningful and meaningless instances are.

conceptual representations; proximal (P) intentions with a mixture of conceptual and perceptual representations; and motor (M) intentions with analog-format representations.

We do not want commit to the specifics of Pacherie's proposal, but we think that her points about (i) the different levels at which the comparator system works, and (ii) the different kinds of representations accessed at each level, are both sensible points. It is at least sensible to think that a monitoring system such as the comparator system has to allow for multiple levels of control. Subjects have to track not only how motor commands are executed, but also whether the intentions that triggered such motor commands are being realized as expected and predicted. Now, we can apply this kind of model to speech generation in general, where the action of speaking begins with an intention (which would be the D-intention) to express a certain thought and culminates with the production of a string of sounds. Speech-related intentions at the different levels generate predictions via the forward model system, which are used to check whether the speech action is being properly realized.

A hypothesis suggests itself at this point: the predictions linked to prior intentions may be made conscious in the same way that we can presumably make conscious the predictions linked to motor commands. Unless we accept a ban on making non-sensory predictions conscious, there is apparently no reason to suppose that we could not make this kind of prediction conscious. Carruthers holds that predictions (sensory predictions, in his case) are made conscious by focusing our attention on them. In general, Carruthers (like Prinz, 2012) believes that consciousness requires attention. There are other hypotheses, though. Jeannerod (1995), for instance, claimed that predictions are conscious just by being predictions of aborted actions, i.e., if an action is aborted after the prediction is issued, the prediction will make into consciousness. His argument is that, when a motor command is aborted, "the motor memories are not or incompletely erased, and the representational levels are kept activated: this persisting activation would thus be the substrate for (conscious) motor images" (Jeannerod, 1995, p. 1429). In any case, our suggestion is that the mechanism that makes sensory predictions conscious may also work for non-sensory predictions.

If this were true, then we may claim that what is made conscious in IS is not just phonological representations, but also their meaning. The prior intention in an act of speaking consists in intending to express a certain thought-content. The prediction corresponding to this kind of intention is the semantic content of the utterance: what we predict, and what we monitor, is that a certain thought-content is expressed. If we were able to broadcast this prediction along with the sensory prediction (i.e., the phonological representations), there would be no need for a further binding of contents into sensory predictions. This seems to be allowed by a theory such as that sketched by Jeannerod (1995), where predictions are conscious by default, but it is more problematic if we follow Carruthers's idea that consciousness requires attention. The trouble in this case is that to be conscious of meaningful IS we would need to attend to two kinds of predictions simultaneously: a prediction about a content, and a prediction about some sounds. In our discussion of Prinz's view in Section "How Thought-Contents are Available to Consciousness,"

we argued that this kind of scenario is not feasible. Yet, we suggest that it is possible to direct our attention not to this or to that particular prediction, but to the outputs of the forward systems (i.e., what the forward systems deliver) considered as a whole. After all, the predictions corresponding to the different layers of intentions are simultaneously active, given that all of them are used in monitoring both the eventual incoming signal and the predictions lower in the hierarchy. This means that the outputs of the forward systems—the cascade of predictions of different levels—form a close network or integrated whole[19].

THE RELATION BETWEEN INNER SPEECH AND UNSYMBOLIZED THINKING

The explanation we just have just outlined has the interesting consequence of allowing us to think about UT in terms of IS without collapsing the former into the latter. In contrast with the format view, the activity view can easily accommodate UT, as this view does not require that a certain format be used for thinking consciously (see Jorba and Vicente, 2014). This is another advantage of the activity view, namely, that by seeing IS as, simply, internal speech, it is not committed to any claim concerning whether or not conscious thinking and phenomenology are possible without a perceptual/sensory medium. However, here we want to move a step further and propose a speculative, though we think plausible, explanation of what UT may be which makes it continuous with IS and begins to account for why we feel authorship with respect to our conscious, but unsymbolized, thoughts (like the judgment that my friend is driving a car).

We just said that it is reasonable to think that the forward system also generates predictions about the likely contents of an utterance. Maybe, we have speculated, this kind of prediction can also be made conscious. Suppose now that we abort a speech action before orders go downstream to motor commands. Then we might get a broadcast prediction about the content of the utterance, which would be experienced as a thought (since it is composed by conceptual/meaning representations). Moreover, there is some chance that it would be experienced as an action because it engages the forward system. At least, minimally, an unsymbolized thought under this construal would feel as initiated (will have the feeling of initiation), as there is an intention in its etiology—which, plausibly, would not be there if we construe UT as simply thoughts (apparently, a thought is not produced by the intention to have it). But it is possible to hold that it would be felt also as authored. As we explained in Section "Is Inner Speech a Prediction?", it is typically said that the sense of agency requires successful comparisons, usually between sensory predictions and sensory signals. But perhaps the comparison between a goal state and a high-level prediction is enough to generate a feeling of agency. Even if not much is known about how the sense of agency is generated in the mental realm (Frith, 2012), we think the possibility that mental agency is related to comparing high level "products" is worth considering.

---

[19]One might contend that Prinz's account can resort to this suggestion, i.e., people may attend to both the acoustic and semantic properties of a sensory representation. However, this suggestion does not help Prinz to avoid our criticism of a regress, giving his commitment to accompanying sensory representations.

If we conceded this view, UT would appear as closely related to IS[20]. We think that this fits nicely the phenomenological characterizations of people reporting UT, in which the subjects have no problems in giving a precise verbal, propositional characterization of what they were thinking yet resist the suggestion that they were experiencing those contents verbally. This easiness of propositional report makes sense if UT is roughly the beginning of a speech act that never became verbally realized. Moreover, the account also advocates a continuity that goes from UT to private speech. Taking into account Vygotsky-inspired approaches, it is not advisable to separate private speech from what we usually call IS, or even from UT, so we see this as a further advantage of our way of looking at IS. The difference between, say, typical IS and muttering, or even private speech, is not a difference in functionality: muttering serves the same general functions as IS (motivation, focusing attention, self-evaluation, etc.). The difference lies in that in typical IS we allegedly produce a prediction about phonological acoustic representations whereas in muttering and in private speech we produce actual sounds. In muttering and private speech, besides, we engage articulation more clearly. In contrast, according to our proposal, in UT we do not even reach the phonological level. Vygotsky claimed that IS is typically condensed with respect to outer speech, and that it is possible for adults to push this condensation to its limit, being able to think "in pure meanings" (see Fernyhough, 2004 for a model of how condensation would proceed). The account here presented would give flesh to this intuition, even though this point of contact with Vygotsky should be regarded as a coincidence (and there are many points of departure from the Vygotskyan tradition: to begin with, UT would not be IS hyper-condensed, but IS aborted before intentions get precise enough). Whether we use one kind of IS, including UT, or the other may depend on stress, the level of attention required, and so on, as Vygotskyans have long claimed[21].

CONCLUSION

We have distinguished two general approaches to the phenomenon of IS: the format and the activity view. The format view,

as endorsed by authors such as Jackendoff, Prinz, and Bermúdez, among others, holds that in IS we recruit a certain format in order to bring thoughts to consciousness. These authors, as well as others who are not particularly interested in the cognitive functions of IS, think about IS as a product, namely, the strings of phonological representations we seem to experience when we talk to ourselves. We have criticized this position on several grounds: first, it has to deny the possibility of conscious UT; secondly, it does not have a clear account as to how thought-contents make it into access-consciousness; and thirdly, it has too narrow a view about the uses of IS. The format view can be weakened in some dimensions, but some problems remain. UT and the agentive experience attached to it remain unexplained, and the issue of how IS makes thoughts conscious is not improved. On top of these general problems, the hypothesis, endorsed by some authors, that IS-as-a-product is a prediction about sensory stimuli, has problems of its own: it is difficult to explain how we can discover errors in our IS if IS is a prediction, and this construal of IS seems incompatible with the idea that alien voices and/or thought insertion are misattributed IS: misattribution seems to require comparison, and a prediction cannot be compared with itself.

Our general diagnosis about the source of all these problems is that supporters of the format view have a narrow focus on issues such as what is constitutive of IS, what is its main function, or what sort of process may be responsible for its production. We have presented an alternative we have labeled "the activity view," which takes a more inclusive view on the IS phenomenon. Describing IS as an activity, namely, speaking, amounts to saying that IS is functionally continuous with overt, or outer, speech. We do not recruit a format with some cognitive purpose, but we speak to ourselves in most of the kinds of situations we speak to other people (self-expression, motivation, attention-focusing, behavior-control, having fun, making irrelevant comments...). This description of what we do in IS suggests that we should think about IS not merely as the output of the linguistic production system, but as the whole action of speaking. Speaking is an action that begins with a prior intention to express a certain thought and plausibly finishes with the production of some sounds that have a certain meaning. The typical IS is that kind of action, except that sounds are not produced but simulated. Adopting this more inclusive view on the phenomenon allows us to solve the problems that affect the format view. First of all, thinking about IS as simply speaking does not question the possibility of UT. Secondly, the view has no problem with explaining the conscious access to thought contents. As it allows that we can think consciously without IS, it is compatible with the view that IS is used only as an aid in some circumstances, lending support to other cognitive functions (e.g., focusing attention in a complex task), or prompting further cognitive resources. Finally, the activity view is in good part motivated by the different uses of IS we can discover.

However, in this paper we have explored other explanatory possibilities for the activity view with several objectives in mind: to be able to capture the intuitive idea that IS proper has meaning, to explain how this meaning can be attached to, and made conscious together with, phonological representations, and to address

---

[20]Following what we said in footnote 8, the hypothesis about how UT is generated we are outlining would link it to the cultural line of development by relating it to IS generation. Yet we do not mean to suggest that UT would be impossible if not related to IS. The explanation we put forward about UT could perhaps be extended to the use of any kind of imagery, although it is not clear to us whether purely imagistic thinking can be propositional. Perhaps our account would predict that non-linguistic creatures could not experience UT, as it is usually characterized.

[21]Another interesting consequence of this view is related to something we mentioned in Section "Is Inner Speech a Prediction?". We said that we are sensitive to mistakes in IS (Oppenheim, 2013), which is problematic for the view that IS is a prediction. In our proposal, which contemplates several levels of predictions and monitoring mechanisms, errors could be detected at the level of motor predictions, especially when these, once they are conscious, re-enter the system as inputs. A prediction cannot check itself, but a higher-order prediction can monitor a low-level prediction and detect errors, even more so, we suspect, if the low-level prediction is also treated as an input for the system. We think that the problems we mentioned in that section are motivated by focusing too narrowly on the motor part of the act of speaking.

two particularly intriguing problems: the nature of UT and the sense of agency attached to it. The proposal we have presented makes use of the characterization of IS as an action in order to explain the binding problem, the nature of UT, and the sense of agency related to conscious thinking. Concerning the binding problem, we have suggested that individuating IS as an action, which begins with a prior intention to express a certain thought, makes it easier to explain how thought-contents are bound into strings of phonological representations. Prior intentions result in predictions about the content of a thought: if such predictions can be made conscious, we have a conscious thought. If the predictions are made conscious together with predictions about phonological representations we have the typical IS ("the little voice in the head"). If the predictions are made conscious alone because the action is aborted very early on, then we have UT. The feeling of agency in this latter case comes from being a cognitive process that is intended, and, plausibly, monitored.

Finally, although we have not tackled the issue of thought insertion in this paper, we think that this general approach is in an overall better position to explain how thoughts may feel as alien, in a way that is parallel to the detection of errors in IS. Higher-level predictions are used to check the correctness of lower level ones in order to monitor whether higher-level intentions are properly realized. Mismatches may result in misattribution and/or error detection. We regard this idea as material for further research.

ACKNOWLEDGMENTS

REFERENCES

Berk, L. E. (1992). "Children's private speech: an overview of theory and the status of research," in *Private Speech: From Social Interaction to Self-Regulation*, eds R. M. Díaz and L. E. Berk (Hillsdale, NJ: Erlbaum), 17–43.

Bermúdez, J. L. (2003). *Thinking without Words*. Oxford: Oxford University Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188

Carpendale, J., Lewis, C., Susswein, N., and Lunn, J. (2009). "Talking and thinking: the role of speech in social understanding," in *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation*, eds A. Winsler, C. Fernyhough, and I. Montero (Cambridge: Cambridge University Press), 83–94.

Carruthers, P. (2009). Mindreading underlies metacognition. *Behav. Brain Sci.* 32, 164–176. doi: 10.1017/S0140525X09000831

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-knowledge*. New York: Oxford University Press.

Carruthers, P. (2014). On central cognition. *Philos. Stud.* 170, 143–162. doi: 10.1007/s11098-013-0171-1

Clark, A. (1998). "Magic words: how language augments human computation," in *Language and Thought: Interdisciplinary Themes*, eds P. Carruthers and J. Boucher (Cambridge: Cambridge University Press), 162–183.

Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas Psychol.* 22, 49–68. doi: 10.1016/j.newideapsych.2004.09.001

Fernyhough, C. (2009). "Dialogic thinking," in *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation*, eds A. Winsler, C. Fernyhough, and I. Montero (Cambridge: Cambridge University Press), 42–52.

Ford, J. M., and Mathalon, D. H. (2004). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *Int. J. Psychophysiol.* 58, 179–189. doi: 10.1016/j.ijpsycho.2005.01.014

Frankish, K. (2010). Evolving the linguistic mind. *Linguist. Philos. Investig.* 9, 206–214.

Frith, C. (2012). Explaining delusions of control: the comparator model 20 years on. *Conscious. Cogn.* 21, 52–54. doi: 10.1016/j.concog.2011.06.010

Geva, S., Bennett, S., Warburton, E. A., and Patterson, K. (2011). Discrepancy between inner and overt speech: implications for post-stroke aphasia and normal language processing. *Aphasiology* 25, 323–343. doi: 10.1080/02687038.2010.511236

Guerrero, M. C. M. (2005). "Methodology of research on inner speech," *Inner Speech—L2: Thinking Words in a Second Language*, ed. M. De Guerrero (New York: Springer), 89–118.

Hatzigeorgiadis, A., Zourbanos, N., Galanis, E., and Theodorakis, Y. (2011). Self-talk and sports performance: a meta-analysis. *Perspect. Psychol. Sci.* 6, 348. doi: 10.1177/1745691611413136

Heavey, C. L., and Hurlburt, R. T. (2008). The phenomena of inner experience. *Conscious. Cogn.* 17, 798–810. doi: 10.1016/j.concog.2007.12.006

Hurlburt, R. T., and Akhter, S. A. (2008). Unsymbolized thinking. *Conscious. Cogn.* 17, 1364–1374. doi: 10.1016/j.concog.2008.03.021

Hurlburt, R. T., Heavey, C. L., and Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Conscious. Cogn.* 22, 1477–1494. doi: 10.1016/j.concog.2013.10.003

Jackendoff, R. (1996). How language helps us think. *Pragmat. Cogn.* 4, 1–35. doi: 10.1075/pc.4.1.03jac

Jackendoff, R. (2007). *Language, Consciousness and Culture: Essays on Mental Structure*. Cambridge, MA: MIT Press.

Jackendoff, R. (2012). *A User's Guide to Thought and Meaning*. New York: Oxford University Press.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia* 33, 1419–1432. doi: 10.1016/0028-3932(95)00073-C

Jeannerod, M. (2006). *Motor Cognition: What Actions Tell the Self*. Oxford: Oxford University Press.

Jorba, M., and Vicente, A. (2014). Cognitive phenomenology, access to contents, and inner speech. *J. Conscious. Stud.* 21, 74–99.

Kozulin, A. (1986). The concept of activity in Soviet psychology: Vygotsky, his disciples and critics. *Am. Psychol.* 41, 264–274. doi: 10.1037/0003-066X.41.3.264

Langdon, R., Jones, S. R., Connaughton, E. and Fernyhough, C. (2009). The phenomenology of inner speech: comparison of schizophrenia patients with auditory verbal hallucinations and healthy controls. *Psychol. Med.* 39, 655–663. doi: 10.1017/S0033291708003978

Langland-Hassan, P. (2008). Fractured phenomenologies: thought insertion, inner speech, thought insertion, and the puzzle of extraneity. *Mind Lang.* 23, 369–401. doi: 10.1111/j.1468-0017.2008.00348.x

Langland-Hassan, P. (2014). Inner speech and metacognition: in search of a connection. *Mind Lang.* 29, 511–533. doi: 10.1111/mila.12064

McCarthy-Jones, S. (2012). *Hearing Voices: The Histories, Causes and Meanings of Auditory Verbal Hallucinations*. Cambridge, MA: Cambridge University Press.

Martínez-Manrique, F., and Vicente, A. (2010). What the. . .! The role of inner speech in conscious thought. *J. Conscious. Stud.* 17, 141–167.

Morin, A. (2011). Self-awareness Part 2: neuroanatomy and importance of inner speech. *Soc. Personal. Psychol. Compass* 2, 1004–1012. doi: 10.1111/j.1751-9004.2011.00410.x

Morin, A., Uttl, B., and Hamper, B. (2011). Self-reported frequency, content and functions of inner speech. *Procedia Soc. Behav. Sci.* 30, 1714–1718. doi: 10.1016/j.sbspro.2011.10.331

Oppenheim, G. M. (2013). Inner speech as a forward model? *Behav. Brain Sci.* 36, 369–370. doi: 10.1017/S0140525X12002798

Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition* 107, 179–217. doi: 10.1016/j.cognition.2007.09.003

Pickering, M., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495

Pitt, D. (2004). The phenomenology of cognition, or, what is it like to think that P? *Philos. Phenomenol. Res.* 69, 1–36. doi: 10.1111/j.1933-1592.2004.tb00382.x

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Prinz, J. (2011). "The sensory basis of cognitive phenomenology," in *Cognitive Phenomenology*, eds T. Bayne and M. Montague (Oxford: Oxford University Press), 174–196.

Prinz, J. (2012). *The Conscious Brain*. New York: Oxford University Press.

Siewert, C. (1998). *The Significance of Consciousness*. New Jersey: Princeton University Press.

Vicente, A. (2014). The comparator account on thought insertion, alien voices and inner speech: some open questions. *Phenomenol. Cogn. Sci.* 13, 335–353. doi: 10.1007/s11097-013-9303-5

Vicente, A., and Martínez-Manrique, F. (2005). Semantic underdetermination and the cognitive uses of language. *Mind Lang.* 20, 537–558. doi: 10.1111/j.0268-1064.2005.00299.x

Vicente, A., and Martínez-Manrique, F. (2008). Thought, language and the argument from explicitness. *Metaphilosophy* 39, 381–401. doi: 10.1111/j.1467-9973.2008.00545.x

Vigliocco, G., and Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychol. Bull.* 128, 442–472. doi: 10.1037/0033-2909.128.3.442

Vygotsky, L. S. (1987). *Thought and Language*. Cambridge, MA: MIT Press.

Winsler, A. (2009). "Still talking to ourselves after all these years," in *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation*, eds A. Winsler, C. Fernyhough, and I. Montero (Cambridge: Cambridge University Press), 3–41.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Inner speech deficits in people with aphasia

Peter Langland-Hassan[1]*, Frank R. Faries[1], Michael J. Richardson[2] and Aimee Dietz[3]

[1] *Department of Philosophy, University of Cincinnati, Cincinnati, OH, USA,* [2] *Department of Psychology, University of Cincinnati, Cincinnati, OH, USA,* [3] *Department of Communication Sciences and Disorders, University of Cincinnati, Cincinnati, OH, USA*

Despite the ubiquity of inner speech in our mental lives, methods for objectively assessing inner speech capacities remain underdeveloped. The most common means of assessing inner speech is to present participants with tasks requiring them to silently judge whether two words rhyme. We developed a version of this task to assess the inner speech of a population of patients with aphasia and corresponding language production deficits. Patients' performance on the silent rhyming task was severely impaired relative to controls. Patients' performance on this task did not, however, correlate with their performance on a variety of other standard tests of overt language and rhyming abilities. In particular, patients who were generally unimpaired in their abilities to overtly name objects during confrontation naming tasks, and who could reliably judge when two words spoken to them rhymed, were still severely impaired (relative to controls) at completing the silent rhyme task. A variety of explanations for these results are considered, as a means to critically reflecting on the relations among inner speech, outer speech, and silent rhyme judgments more generally.

Keywords: inner speech, aphasia, subvocalization, rhyming, attention, executive function, stroke

Introduction

Inner speech is the little voice in the head, sometimes known as thinking in words. It is the capacity to say things to oneself, silently. When people are asked at random and unexpected intervals to report on the nature of their current conscious experience, they report being engaged in inner speech 20% of the time, on average (Heavey and Hurlburt, 2008).

There are many current proposals concerning the role of inner speech in cognition. One account holds that inner speech is an important element of working memory, with inner speech utterances consisting in recitations within a "phonological loop," which serves to keep limited amounts of information readily at hand to multiple processing units (Baddeley, 2007). More recently, researchers have found evidence that inner speech underlies certain executive functions, such as the ability to switch between cognitive tasks (Emerson and Miyake, 2003; Vygotsky et al., 2012/1962) and to engage in flexible problem solving of the kind required by the Wisconsin Card Sorting Task (Baldo et al., 2005). A role for inner speech has also been proposed in reading. When people are given information about a speaker's accent and speaking rate, this influences the rate at which they read text that putatively records that person's speech (Alexander and Nygaard, 2008; Kurby et al., 2009).

Other views locate inner speech even more centrally within the mind, proposing that a certain kind of thinking is essentially dependent on inner speech. This viewpoint has a long history in psychology, represented by Watson (1930), Paivio (1971, 1986) and Vygotsky et al. (2012/1962).

More recently, Carruthers (2002) has argued that inner speech allows us to collect and integrate information from a variety of cognitive modules; and Gauker (2011) has proposed that all thought that deserves to be called *conceptual thought* occurs in inner speech. A number of theorists have also proposed a role for inner speech in metacognition (thinking about one's own thinking) and self-awareness (Jackendoff, 1996; Clark, 1998; Bermudez, 2003; Morin, 2009; Carruthers, 2011).

One of the most effective ways to assess proposals concerning the role of inner speech in cognition is to compare the abilities of people with impaired or absent inner speech to those whose inner speech is intact. But how can we know whether, and to what degree, someone is able to generate inner speech? How can we assess, in a scientifically rigorous way, whether someone has the ability to say things to him or herself silently? Historically, the primary means for such an evaluation has been to present participants with pairs of words and to ask them to judge, silently, whether the words rhyme (e.g., "chair" and "care"), or whether they are homophones (e.g., "stare" and "stair"; Levine et al., 1982; Feinberg et al., 1986; Geva et al., 2011a,b). Such studies use a high proportion of target words-pairs that rhyme but which do not have similar endings (e.g., "box" and "socks"), to prevent participants from answering based on the words' visually apparent orthography. Occasionally non-words are used as stimuli (e.g., "pole" and "voal") to prevent participants from answering based on knowledge of orthography (see, e.g., Geva et al., 2011a,b). Alternatively, pairs of pictures may be used as stimuli, with participants being asked to indicate whether the words for the pictured objects rhyme (Sasisekaran et al., 2006). This approach reduces possible interference by difficulties a participant may have with reading or with language reception generally.

Intuitively, we make silent rhyme or homophone judgments by uttering the relevant words in inner speech. However, inner speech cannot simply be defined operationally as the ability to silently make correct judgments about the sounds of words. After all, a person who accurately *guesses* whether a pair of words rhyme would not in virtue of that success count as having used inner speech. And, moreover, it is possible that a person may be able to silently utter words in inner speech yet nevertheless be unable to accurately judge whether they rhyme. However, there are at present no better-established or more reliable means for assessing inner speech abilities.

The present study seeks to better understand the relationship between inner speech and the ability to silently judge rhymes, by assessing the relationship between outer speech and silent rhyme judgments in a population with post-stroke aphasia. People with aphasia (PWA) have impaired language capacities due to one or more neural lesions, typically acquired as a result of stroke. Depending on the location of the lesion, a patient's deficits may center more on the production of speech (as in Broca's, conduction, and anomic aphasia), or on the comprehension of speech (as in Wernicke's aphasia). However, almost all PWA have at least some difficulties with respect to both speech production and comprehension.

People with aphasia have been shown in many cases to have impaired inner speech (Levine et al., 1982; Feinberg et al., 1986;

Geva et al., 2011a,b). Yet there is relatively little data available concerning the precise relationship between outer speech and inner speech abilities in PWA. Nor, for that matter, is there much data available on the relation between inner and outer speech abilities in the general population. Geva et al. (2011a) offer the most thorough examination of the relation between inner and outer speech in PWA. Their work confirms that PWA are in general significantly impaired at silent rhyme and homophone tasks, compared to controls. They note, however, a high degree of variability in inner speech abilities, with some patients performing at normal levels, while others are clearly impaired.

Geva et al. (2011a) also sought to assess the relationship between inner and outer speech abilities in PWA. Their main task required patients to read pairs of words or non-words both silently (in one condition) and out loud (in another). In one condition, they were asked to judge whether the two words rhymed; in a second they were asked to judge whether the two words were homophones; and, in a third, they were asked to judge whether the two non-words were homophones. In each case the judgments of PWA during the silent condition were impaired relative to controls. And while significant correlations were found between the inner and outer speech abilities of PWA, their data also revealed a number of interesting dissociations in individual patients. Some patients showed relatively strong performance in the silent versions of the task, yet were impaired on the overt versions; this suggests intact inner speech with impaired overt speech. And, perhaps more surprisingly, dissociations were also found in the opposite direction, with some participants showing impairments in the silent rhyme judgments, yet normal performance in the overt rhyme judgment tasks. This in turn suggests impaired inner speech in the presence of relatively normal overt speech. While reports of this dissociation are rare, there are precedents (Levine et al., 1982; Papafragou et al., 2008). Moreover, the rarity of such reports may be due in part to the fact that inner speech capacities are seldom assessed in people with relatively intact overt speech.

In light of these somewhat surprising dissociations, Geva et al. (2011a) urge that more data should be collected concerning the inner speech abilities of PWA. The present study, conducted as part of a larger investigation of the relation between inner speech and metacognition (under review), is reported with that in mind. From a therapeutic standpoint, if deficits in overt speech are not always accompanied by deficits in inner speech, this could open the door to therapeutic interventions that make use of a patient's preserved inner speech. Arguably, assessments of inner speech should then become a standard component of aphasia screening tests. And, from a more theoretical perspective, if there are significant dissociations between outer speech capacities and capacities to silently judge rhymes, this may call into question the links many presume between inner speech and outer speech. Specifically, it may challenge a common view of inner speech as overt speech minus a motor component (Oppenheim and Dell, 2010; Carruthers, 2011; Pickering and Garrod, 2013). At the same time, dissociations between outer speech and the ability to silently judge rhymes may lead us to consider more carefully the

possibility that silent rhyme judgments require something more than intact inner speech.

In the present study, we first sought to confirm whether PWA would show deficits compared to controls in a pictorial silent rhyme judgment task that [unlike Geva et al. (2011a)] did not require them to read words. A second interest was in whether the performance of PWA on the silent rhyme task would be correlated with their ability to judge whether words overtly spoken to them rhyme. In this way, we sought to understand the relation between judging rhymes, in general, and judging rhymes silently. A third interest was to investigate what correlations may exist between the silent rhyming abilities of PWA and their abilities on various cognitive and linguistic tests of the Western Aphasia Battery-Revised (WAB-R; Kertesz, 2006) and Cognitive Linguistic Quick Task (CLQT; Helm, 2003), including tests of confrontation naming abilities, generative naming abilities, executive function, and attention. In this way we sought to better understand the relation of the capacity for making silent rhyme judgments to an array of other seemingly closely related abilities.

Materials and Methods

Participants
Patients with Aphasia
A total of 11 individuals (4M/7F, 10 right-handed, one left-handed, mean age 60.3 ± 8.0, age range 44–76, mean years of education 15.3 ± 2.1) with chronic post-stroke aphasia (mean months post-stroke=124 ± 77.9) were recruited from a registry of patients held at the University of Cincinnati College of Allied Health Sciences, Division of Communication Sciences and Disorders[1]. All were native English speakers. A diagnosis of aphasia was confirmed based on their performance on the Western Aphasia Battery – Revised (Kertesz, 2006). All PWA

---

[1]A 12th person with a history of aphasia was recruited and tested, but excluded from analysis because his Aphasia Quotient (=96) on the WAB-R was not sufficiently low to qualify him for a diagnosis of aphasia.

TABLE 1 | Demographic information, participants with aphasia.

	Sex	Age	Years of education	Months post-onset	Aphasia type
201	Female	59	14	72	Conduction
202[1]	Male	44	16	112	Broca's
203	Male	58	18	76	Anomic
204[2]	Female	68	16	175	Broca's
206	Female	76	14	315	Broca's
207	Female	56	14	92	Broca's
208	Female	54	12	134	Anomic
209	Male	60	16	101	Broca's
210	Female	67	18	100	Broca's
211	Female	62	16	172	Conduction
213	Male	59	14	15	Broca's

[1]*Apraxia of Speech present based on clinical judgment.*
[2]*Left handed pre-stroke.*

received a diagnosis of Conduction, Broca's, or Anomic aphasia (see **Table 1**), and all exhibited significant difficulties with language production. Their language comprehension capacities, however, were relatively intact, allowing them to given informed consent and to understand task instructions.

Control Participants
A total of 12 healthy volunteers (4M/8F, mean age 58.7 ± 8.5, age range 47–78, mean years of education 14.8 ± 1.7) were chosen for a control group, roughly matched in age [$t(21) = 0.47$, $p > 0.64$], gender, and education [$t(21) = 0.55$, $p > 0.58$] to the PWA. All were native speakers of English, with no history of stroke or other neurological or psychiatric disorders.

Materials
All PWA were administered basic vision and hearing screening exams, the WAB-R (Kertesz, 2006), and the CLQT (Helm, 2003). An overt rhyme judgment task (described below) was also administered to PWA.

For the silent rhyming task, 88 digital photographs and drawings were used to create 44 trials involving two pictures each. These trials were presented on an Asus 8A-Series computer with a 21-inch touch-sensitive screen. The program was written in C++ and recorded responses and response times automatically (10 ms resolution).

Procedure
In the first session, all PWA completed the WAB-R and CLQT tests, as well as basic hearing and vision screening tests. The WAB-R was used to confirm aphasia severity and type, and to assess overt word production abilities. Of particular interest was participant performance on confrontation naming tasks (where objects are shown to the participant and the participant must name them), and generative naming tasks (where a category, such as "animals," is given to participants and they must name as many members of that category as possible). The CLQT also included confrontation and generative naming tasks, together with a variety of non-linguistic cognitive tests that were used to rule out the possibility that task performance was due to cognitive limitations unrelated to the patients' language deficits.

At the beginning of the second session, the overt rhyme judgment task was administered to PWA. During this task, 10 pairs of one-syllable words were spoken aloud by the experimenter, with patients indicating (Yes or No) whether the words rhymed. The mean word frequency for the words used (using log frequency sore) was 1.75 (SD = 0.51; Medler and Binder, 2005). This task was administered in order to investigate the relation between judging rhymes that are heard and judging rhymes silently, through inner speech.

After completing the overt rhyming task, PWA were administered the silent rhyming task. Controls were tested for only a single session, and were not administered the overt rhyming task, the WAB-R, or the CLQT. Controls were not administered these tests mainly because they were not relevant to the larger study from which this report is retrospectively made.

The silent rhyming task was administered as follows. On a touchscreen computer, participants were shown 44 sets of two

pictures each, one set at a time, and were asked to indicate, without speaking aloud, whether the words for the pictured items rhyme (**Figure 1**). The first four trials were training trials, during which a team of experimenters demonstrated the task by first completing it out loud, and then silently. The experimenters read from a script to ensure that the task was explained in the same way for all participants. During training trials, it was emphasized that participants must answer the questions without making any vocal sounds. In addition to answering yes (by touching a green check) or no (by touching a red X), participants could indicate that they did not know by touching a blue question mark. Touching the blue question mark was counted as an incorrect answer, for purposes of scoring. Among the 40 test trials, 20 presented stimuli that rhymed, while 20 did not. Of the 20 rhyming trials, 10 involved pictures of items whose linguistic labels rhymed but did not share similar orthographic endings (e.g., "box" and "socks"). This was to decrease the likelihood that participants could answer by forming visual images of written words as opposed to using auditory-phonological cues. The mean word frequency rating (using the log frequency score) for the words tested was 1.38 (SD = 0.87; Medler and Binder, 2005)[2].

Results

Silent Rhyming Impairment in PWA

An independent-samples t-test was conducted to compare silent rhyming task performance (i.e., hits) in control and PWA conditions. This analysis revealed a significant difference between controls (mean = 36.4, SD = 3.29) and PWA (mean = 21.64, SD = 4.30); $t(21) = 9.32$, $p < 0.001$. There was also a significant difference between controls and PWA with regard to d-prime, $t(21) = 9.18$, $p < 0.001$, (see **Figure 2**), with PWA essentially guessing when providing their answers on the silent-rhyme task. (d-prime is a sensitivity index from signal detection

---

[2]The word frequency ratings for the silent rhyming words were calculated after removing 'can' from the group. 'Can' was an outlier in having an extremely high frequency rating, due to its common use as an auxiliary verb (The McWord database is only sensitive to orthographic form). The silent rhyming test, however, made use of 'can' in its noun sense, showing a picture of a can—hence its exclusion.

FIGURE 1 | Example of a silent rhyming task trial.

theory which, for the present study, captured the ability of participants to discriminate or detect whether two words rhymed.) Notably, the average number of false alarms (touching the check mark or opting out when there was not a rhyme) for PWA was 10.18 (SD = 4.56), compared to 0.58 (SD = 1.165) for controls [$t(21) = 7.06$, $p < 0.001$]. There was also a similarly large difference between the average number of misses (touching the red X or opting out when there was a rhyme) for PWA (mean = 8.18, SD = 3.92) compared to controls [mean = 3.0, SD = 2.45; $t(21) = 3.84$, $p < 0.01$]. **Table 2** shows the respective means for PWA and controls on the silent rhyming task with respect to hits, correct rejections, false alarms, misses, opt-outs, and d-prime. Finally, the mean number of PWA who correctly identified a rhyming pair as rhyming was not significantly higher when the orthography of the word-endings matched (mean = 7.6; ±1.43) than when the word-endings did not orthographically match [mean = 7.2; ±1.23; $t(9) = -0.684$, $p > 0.51$].

Lack of Correlation Between Silent-Rhyming and Overt Rhyming

On the overt rhyme task, the performance of PWA was considerably better than their performance on the silent rhyme task, with a mean score of 8.67 out of 10 (SD = 1.2). Spearman's Rank order correlations were assessed with respect to the silent rhyming scores of PWA and their scores on the overt rhyme task (see **Table 3**). For PWA, the correlation between silent rhyme performance and overt rhyme performance fell short of significance ($r = 0.443$, $p > 0.17$). Possible reasons for this null effect are discussed below (Controls did not complete the overt rhyme task, for reasons discussed above.).

Lack of Correlations Between Silent Rhyming and Sub-Tests of CLQT and WAB-R

The CLQT and WAB-R tests revealed the following impairments in PWA. On the CLQT, four out of 11 fell below the cut-off score for normal performance (as established and validated by the designers of the CLQT) on sub-tests for Attention (mean = 176.7, SD = 19.5), while six of 11 fell below the cut-off score for normal performance on sub-tests for Executive Functions (mean = 24.5, SD = 6.8). Seven out of 11 fell below normal limits on the CLQT confrontation naming task (mean = 8.2, SD = 2.4), while 11 out of 11 fell below normal limits on the CLQT generative naming task (mean = 1.8, SD = 1.2). The mean score of PWA with respect to object naming on the WAB-R was 40, out of a possible 60 (SD = 17.2), while the mean score of PWA on the WAB-R word fluency test was 6.4, out of a possible 20 (SD = 4.2) (Controls did not complete the WAB-R or CLQT, for reasons discussed above.).

No significant correlations were found between the silent rhyming performance of PWA and the performance of PWA on these sub-tests of the WAB-R and CLQT (all $p > 0.49$; see **Table 4**). For instance, the correlation between PWA silent rhyming performance and the WAB-R Object Naming task was only $r = -0.030$; and the correlation between PWA silent rhyming scores and their CLQT confrontation naming scores was $r = -0.104$. Nor were there significant correlations found

FIGURE 2 | (A) Number of correct responses (hits) on the silent rhyming task for people with aphasia (PWA) and controls; error bars correspond to standard deviations of the mean. **(B)** Box plot of the *d*-prime scores for PWA and controls.

TABLE 2 | Mean scores by population on silent rhyming task.

	People with aphasia (PWA)	Controls
Hits[1]	11.8	17.0
Correct rejections[2]	9.8	19.4
Misses[3]	8.2	3.0
False alarms[4]	10.2	0.6
Opt-out[5]	1.7	0.6
d-prime	0.2	2.7

[1]*Hits = touching green check for rhyming pairs.*
[2]*Correct rejections = touching red X for non-rhyming pairs.*
[3]*Misses = touching red X or opting-out when pair in fact rhymed.*
[4]*False alarms = touching green check or opting out when pair did not rhyme.*
[5]*Opt-out = touching blue question mark.*

between PWA silent rhyming scores and Executive Function ($r = -0.294$) or Attention ($r = -0.242$) scores on the CLQT.

Significant correlations were, however, found between certain subtests of the WAB-R and CLQT (see **Table 4**). For instance, the scores for PWA on the WAB-R object naming were highly correlated ($r = 0.803$) with their scores on the CLQT confrontation naming task. And the scores for PWA on the WAB-R word fluency task were highly correlated ($r = 0.868$) with their scores on the CLQT generative naming task.

Discussion

In this study we sought to assess the degree to which the inner speech of people with known outer speech deficits (due to aphasia) is impaired, relative to controls. We also sought to assess the degree of correlation between their inner speech impairments and their overt speech and rhyming abilities. And, finally, we wanted to investigate what correlations there might be between their inner speech abilities and their aptitude on measures of executive function and attention. In this way, we hoped to gain a clearer understanding of the relation between inner and outer speech, and between inner speech and executive function and attention. Currently, the degree to which inner speech is a distinct

mental capacity, dissociable from these others, is not well understood.

Our main finding was that PWA (with, specifically, Broca's, conduction, and anomic aphasia) have great difficulty completing silent rhyming tasks, compared to controls. Insofar as performance on the silent rhyming task is a reliable indicator of inner speech ability, their inner speech was severely impaired. More surprisingly, however, we did not note any significant correlations between their silent rhyming abilities and their abilities on various overt rhyming and overt speech tasks. We make note of these and other null effects—and discuss them further below—not because any strong inferences can be made from the lack of such correlations, but because they are of interest in considering possibilities for future research. In particular, they raise interesting questions concerning the degree to which inner speech may be a distinct capacity, dissociable both from overt speech and from cognitive capacities such as executive function and attention.

The deficits of PWA on the silent rhyming task compared to controls were even more pronounced than those found by Geva et al. (2011a), whose patients with aphasia answered approximately 80% of silent rhyme and homophone judgment prompts correctly. In the present study, PWA answered only 54% of silent rhyme questions correctly, and had a high proportion of false alarms [mean = 10.18 (±4.56)]. Such results would be expected if participants were simply unable to perform the task and were guessing on each trial.

One might hypothesize that the greater difficulty shown by patients in the present study resulted from their having to find the proper word corresponding to each image before generating the words in inner speech. In the case of Geva et al. (2011a) the words were given to patients in written form, obviating the need to find the proper word for the objects. However, the data recorded on overt naming abilities in this group of PWA does not support this interpretation. For if the relative difficulty of the silent rhyming task was to be explained by a general inability of patients to generate words for the pictured items, we would expect patients to show difficulties *both* with the silent rhyming task and with ordinary confrontation naming tasks. For both kinds of task confront

TABLE 3 | Raw scores of participants with aphasia on rhyming, WAB-R, and CLQT tests.

ID#	Silent rhyming	Overt rhyming	Western Aphasia Battery-Revised (WAB-R) aphasia quotient	WAB-R object naming	WAB-R word fluency	Cognitive Linguistic Quick Task (CLQT) confrontation naming	CLQT generative naming	CLQT attention[5]	CLQT executive functions[6]
Max[1]	40	10	100	60	20	10	Unlimited	215	40
Min[2]	NA[3]	NA	93.8	NA	NA	10	5	180	24
201	17	9	81.8	51	9	9	1	197	37
202	28	9	71.2	55	11	10	3	191	30
203	22	10	68.9	30	7	7	2	189	27
204	26	8	43.9	26	1	6	1	180	20
206[4]	20	9	66.1	47	5	10	2	137	10
207	17	6	50.6	36	7	8	2	199	29
208	27	10	73.4	57	3	10	1	153	22
209	15	7	70.9	50	10	10	3	182	28
210	24	8	30.2	0	0	2	0	170	20
211	21	9	81.1	55	13	9	4	161	23
213	21	9	59.2	33	4	9	1	185	23

[1]Maximum possible scores are indicated for each subtest.
[2]Minimum score for normal performance. Any lower score is considered impaired.
[3]The mean score for controls on this test was 36.4.
[4]Minimum score of normal performance on CLQT is lower for this participant, given age (see note 5).
[5]Ranges for participants up to 69 years old: within normal limits = 215–180, Mild = 179–125, Moderate = 124–50, Severe = 49–0; for participants over 70 years and older: within normal limits: 215–160, Mild: 159–100, Moderate: 99–40, Severe: 39–0.
[6]Ranges for participants up to 69 years old: WNL: 40–24, Mild: 23–20, Moderate: 19–15, Severe: 14–0; for participants 70 years and older: WNL 40–19, Mild: 18–14, Moderate: 13–8, Severe 7–0.

TABLE 4 | Between measure correlations (Spearman's rank order correlation).

	Silent rhyming	d'	Overt rhyming	WAB aphasia quotient	WAB object naming	WAB word fluency	CLQT confrontation naming	CLQT generative naming	CLQT attention	CLQT executive function
SR	–	0.991 (0.000)	0.443 (0.173)	–0.100 (0.769)	–0.030 (0.931)	–0.336 (0.312)	–0.104 (0.761)	–0.232 (0.473)	–0.242 (0.473)	–0.294 (0.381)
d'	0.991 (0.000)	–	0.475 (0.139)	–0.118 (0.729)	–0.018 (0.957)	–0.432 (0.185)	–0.104 (0.762)	–0.323 (0.332)	0.287 (0.392)	–0.348 (0.295)
OLR	0.443 (0.173)	0.475 (0.139)	–	0.527 (0.096)	0.379 (0.250)	0.046 (0.894)	0.275 (0.413)	–0.025 (0.942)	–0.254 (0.451)	–0.067 (0.844)
WAB aphasia quotient	–0.100 (0.769)	–0.118 (0.729)	0.527 (0.096)	–	0.866 (0.001)	0.697 (0.017)	0.606 (0.048)	0.438 (0.178)	0.036 (0.915)	0.507 (0.112)
WAB object naming	–0.030 (0.931)	–0.018 (0.957)	0.379 (0.250)	0.866 (0.001)	–	0.607 (0.048)	0.803 (0.003)	0.488 (0.127)	–0.100 (0.769)	0.373 (0.259)
WAB word fluency	–0.336 (0.312)	–0.432 (0.185)	0.046 (0.894)	0.697 (0.017)	0.607 (0.048)	–	0.461 (0.153)	0.868 (0.001)	0.342 (0.304)	0.687 (0.020)
CLQT confrontation naming	–0.104 (0.761)	–0.104 (0.762)	0.275 (0.413)	0.606 (0.048)	0.803 (0.003)	0.461 (0.153)	–	0.479 (0.136)	–0.216 (0.523)	0.179 (0.598)
CLQT generative naming	–0.232 (0.493)	–0.323 (0.332)	–0.025 (0.942)	0.438 (0.178)	0.488 (0.127)	0.868 (0.001)	0.479 (0.136)	–	0.071 (0.837)	0.355 (0.284)
CLQT attention	–0.242 (0.473)	0.287 (0.392)	–0.254 (0.451)	0.036 (0.915)	–0.100 (0.769)	0.342 (0.304)	–0.216 (0.523)	0.071 (0.837)	–	0.854 (0.001)
CLQT executive function	–0.294 (0.381)	–0.348 (0.295)	–0.067 (0.844)	0.507 (0.112)	0.373 (0.259)	0.687 (0.020)	0.179 (0.598)	0.355 (0.284)	0.854 (0.001)	–

the participant with objects (or pictures of objects) and require that they name them, either in inner speech, or overtly. Yet, as evidenced by the scores of PWA on the WAB object naming task and the CLQT confrontation naming task (both of which are confrontation naming tasks) many of the patients showed relatively intact confrontation naming abilities, with several performing within normal limits (see **Table 3**). So the severity of the silent rhyme deficits observed, compared to Geva et al. (2011a), might not be due to general difficulties generating words for presented objects. Indeed, while the PWA showed a broad continuum of abilities on the confrontation naming tasks, there was not a significant correlation between those abilities and their scores on the silent rhyming task. Yet one would expect there to be such a correlation if an inability to succeed at confrontation naming (both inner and overt) explained their difficulties. Moreover, the mean word frequency ratings for the words featured in the CLQT object naming task (mean = 1.36; ±0.57) and WAB confrontation naming task (mean = 1.26; ± 0.70) are lower than for those featured in the silent rhyming task (mean = 1.38; ± 0.87). The better performance of PWA on the overt naming tasks than silent rhyming task is therefore not a result of the silent rhyme task using less familiar words[3].

If a general inability to name objects does not explain the poor performance of PWA on the silent rhyming task, what does? One possibility is that, while the PWA were able to generate the words for the pictured items in inner speech, they were unable to reliably judge whether the words rhyme, due to a specific deficit in discriminating rhyming words from non-rhyming words. However, the data do not support this interpretation. The abilities of PWA to judge rhymes when word pairs were spoken to them by the experimenter was relatively intact, with 87% of their answers being correct, as compared to only 54% for the silent rhyming task. It should be noted, however, that the mean word frequency of the words used in the overt rhyming task (1.75, ±0.51) was significantly higher [$t(96) = 2.55$, $p = 0.012$] than that of the words used in the silent rhyming task (1.38, ±0.87). It is possible that the increased familiarity of the words spoken aloud to participants played some role in facilitating their ability to judge whether the words rhymed. Nevertheless, it is less obvious in this case, as compared to the confrontation naming tasks, why word frequency would influence performance. Participants did not, after all, have to generate the relevant words for the overt rhyming task; they only had to listen to the words and judge whether they rhymed. Nor was there a significant correlation observed between overt rhyming scores and silent rhyming scores. Indeed, eight out of the 11 patients had scores of either 9 or 10 (out of 10) on the overt rhyming task (**Table 3**); yet the mean score among those same eight patients on the silent rhyming task was 22.3 (out of 40), or only 56% correct. Thus, it is not clear that the poor performance exhibited by PWA on the silent rhyming task, compared to controls, can be explained by a general impairment in judging aurally presented rhymes.

---

[3]A reviewer raises the possibility that the better performance of patients on silent rhyming tasks in Geva et al. (2011a) compared to the present study may have been due to the potentially greater word-frequency of the words used by Geva et al. (2011a). Geva et al. (2011a) do not report the word frequency ratings for their stimuli, nor the specific words used, so we cannot assess this possibility.

That said, it is worth bearing in mind the $r = 0.443$ correlation found between silent rhyming scores and overt rhyming scores, for the PWA (**Table 4**). While not statistically significant, this degree of correlation in our low-powered sample warrants further investigation of the link between these two abilities. A possibility worth bearing in mind is that even judging aurally perceived rhymes requires some degree of instantaneous "replay" of the words in inner speech. In that case, an inability to judge whether aurally perceived word pairs rhyme could be explained in terms of an inability to generate words in inner speech, and not vice versa.

Nevertheless, in our sample, patients could reliably judge whether words spoken to them rhymed, and were in many cases relatively unimpaired at overtly naming pictured objects, yet were without exception unable to complete the silent rhyme task comparably to controls[4]. There are at least two ways to interpret this finding. First, it could be that the preserved ability of some patients to *overtly* name objects with which they are confronted was not matched by a comparable ability to generate the names for objects *in inner speech*. This would be a surprising finding in light of theories that conceive of inner speech as motor-precursor to outer speech (Oppenheim and Dell, 2008; Pickering and Garrod, 2013). Yet it may be less surprising from a Vygotskian perspective, which views inner speech as an internalized, and developmentally posterior, version of outer speech (Jones and Fernyhough, 2007; Vygotsky et al., 2012/1962). The hypothesis that inner speech deficits were not matched by comparable overt naming deficits meshes with a handful of other studies that have reported intact language in the absence of inner speech (Levine et al., 1982; Papafragou et al., 2008; Geva et al., 2011a), and with a recent study showing distinct neural correlates for inner and outer speech (Geva et al., 2011b).

An alternative hypothesis is that, while the PWA had no deficits in using inner speech to inwardly name the pictured objects (compared to their overt confrontation naming abilities), and no severe impairments judging rhymes when they were heard, nevertheless the task as a whole presented a cognitive load that was too great to overcome, given their impairments. That is, perhaps silently judging rhymes in inner speech requires working memory resources, or executive function abilities, that PWA lack. To be clear, this hypothesis holds that the specific resource lacked is not that which makes it possible to utter the relevant words in inner speech, or to judge rhymes in general. The idea is that it may be something else, such as the resource that makes it possible to hold two words in mind long enough to judge whether they rhyme (i.e., working memory), or that which allows one to assign requisite attention to the task (e.g., executive function). It should be noted, however, that the PWA did not in general show significant deficits in executive functions or attention, on the CLQT. With respect to executive functions, five out of 11 scored within normal limits, with five having only mild impairments, and one having moderate impairments (see **Table 3**). And, with respect to attention, seven of 11 scored within normal limits, with the other four being only mildly impaired. Nor were there

any significant correlations between silent rhyming scores and executive function or attention scores on the CLQT. Moreover, it should be noted that some of both the attention and executive functions sub-tests explicitly required language use, which was of course known to be impaired in these participants. Their relatively strong cumulative scores for executive functions and attention therefore suggest that they did not have considerable cognitive impairments *outside of* their specific linguistic deficits. It is therefore unclear what sort of general, non-linguistic cognitive deficit might have accounted for the special difficulties PWA had, compared to controls, on the silent rhyming task. Another possibility is that the CLQT is not sensitive to the reduced ability to allocate attention to language-related tasks because it relies largely on non-verbal cognitive tasks. It is well accepted in aphasiology that aphasia can be explained as a deficit in resource allocation, specifically for language (while other aspects of cognition are intact; Murray, 1999). For now it remains possible that the capacity to generate inner speech is simply a distinct ability from executive function, attention, outward rhyme judging, and overt naming—one with its own neural substrate, and which can be severely impaired without comparable impairments in these other capacities.

Future work could further tease apart these hypotheses by using other implicit measures of inner speech on a similar population. For instance, studying populations whose native languages encode information about bounded motion differently, Papafragou et al. (2008) found that the way a speaker's native language encodes such information influences their eye movements when surveying an action sequence that they are told to remember. This suggests that inner speech may be influencing visual search in such cases. If the eye movements of PWA did not, under such conditions, match the patterns typical of controls with the same native language, this would be corroborating evidence that they did not in fact have intact inner speech (even if their outer speech was comparably preserved). It bears noting, however, that language-learning may affect thought by influencing the way in which non-linguistic thought structures develop, and not necessarily through the mediation of inner speech.

Whatever the explanation for the low scores of PWA on the silent rhyming task, it is certainly of interest to note that no significant correlations were found between the various language and cognitive assessment scores patients received on the WAB-R and CLQT, and their scores on the silent rhyming task. By contrast, patient scores on the WAB and CLQT tests were often highly correlated with each other, as one would expect (see Results; **Table 4**). One reason for the lack of correlations may simply be due to the relatively small sample size (i.e., low power). An additional explanation for the lack of correlations may be that all, or almost all, of the PWA were simply incapable of completing the silent rhyme task. The highest score of any patient on the task was 28 (out of 40), which was over 2.5 SDs below the mean for controls; and the mean score for PWA was barely above 50% correct. Furthermore, the mean *d*-prime score for PWA on the silent rhyming task was only 0.2, which is little better than what would be expected (0.0) if they were all guessing all of the time. Thus, their ability, as a group, to discriminate rhyming trials from non-rhyming trials

---

[4]The highest score of any patient on the silent rhyme task was over 2.5 SD below the mean for controls on that task.

was sufficiently low to render unlikely any meaningful correlations with their more widely distributed scores on the WAB-R or CLQT.

Left unanswered, however, is why the patients whose aphasia was mild by comparison to others and who scored at normal or near-normal levels on some language production tasks (e.g., 201, 202, 211) did not have comparably better scores on the silent rhyming task. A further datum worth noting in this regard is that our population of PWA all had significant deficits on the generative naming components of the WAB-R and CLQT (The WAB-R generative naming task is called "Word Fluency," and the CLQT generative naming task is called "Generative Naming" on **Table 3**.). Only one participant (211) had scores on either test that approached levels typical for controls. Thus we did not observe the same kind of dissociation between generative naming abilities and silent rhyming abilities as we did between confrontation naming and silent rhyming. It is not immediately obvious why severe impairments in silent rhyming would co-occur with severe impairments in generative naming, as opposed to confrontation naming.

One possible explanation is that, on the CLQT and WAB-R confrontation naming tasks, participants were not in the situation of having to choose which of several common names to give for the pictured objects. This is because the CLQT counts any commonly used name that is produced for a pictured item as a correct response, while the WAB-R is designed so as to only feature stimuli with one common name. By comparison, the silent rhyming task has some of the character of a generative naming task, to the extent that participants may have had to audition (and therefore generate) multiple appropriate names for a single object in order to identify whether each trial was a rhyming trial. When faced, for instance, with a picture of a box and a pair of socks, a participant's success may have required moving past a first word generated in response to the stimulus (e.g., "package"), to find another (e.g., "box") that rhymed with the word for the companion picture (e.g., "socks"). If this were the case, it would help explain why serious deficits in generative naming went hand-in-hand with troubles on the silent rhyming task[5]. That said, there were no statistically significant correlations between performance on the generative naming tasks and performance on the silent rhyming task. It would be useful, in future work, to investigate whether

clearer dissociations or correlations can be found between silent rhyming abilities and generative naming abilities, by ensuring that the stimuli used in silent rhyming tasks do not require or encourage participants to potentially generate multiple words in response to the stimulus. One way to do this would be to use pairs of written words as stimuli, as in Geva et al. (2011a,b).

Conclusion

Our participants with aphasia showed severe deficits at the kinds of silent rhyming tasks that are typically used to assess inner speech abilities. This suggests that PWA with impaired overt speech typically have impaired inner speech as well. Interestingly, even patients with relatively preserved confrontation naming and overt rhyming abilities performed at chance on the silent rhyming task. This highlights the possibility that inner speech abilities are often more severely impaired in PWA than overt speech capacities. While more research must be done before any strong conclusions can be made, it may simply be that generating and using inner speech is more cognitively and linguistically demanding than generating overt speech and, further, that the neural substrates for each are somewhat distinct (Geva et al., 2011b). This would mesh well with a Vygotskian perspective on which inner speech develops posterior to overt speech, and is by comparison a more sophisticated and cognitively demanding activity.

---

[5] We thank a reviewer for raising this possibility.

References

Alexander, J. D., and Nygaard, L. C. (2008). Reading voices and hearing text: talker-specific auditory imagery in reading. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 446–459. doi: 10.1037/0096-1523.34.2.446

Baddeley, A. D. (2007). *Working Memory, Thought and Action*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198528012.001.0001

Baldo, J. V., Dronkers, N. F., Wilkins, D., Ludy, C., Raskin, P., and Kim, J. (2005). Is problem solving dependent on language? *Brain Lang.* 92, 240–250. doi: 10.1016/j.bandl.2004.06.103

Bermudez, J. L. (2003). *Thinking Without Words*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195159691.001.0001

Carruthers, P. (2002). The cognitive functions of language. *Behav. Brain Sci.* 25, 657–726. doi: 10.1017/S0140525X02000122

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199596195.001.0001

Clark, A. (1998). "Magic words: how language augments human computation," in *Language and Thought: Interdisciplinary Themes*, eds P. Carruthers and J. Boucher (Cambridge: Cambridge University Press), 162–183.

Emerson, M. J., and Miyake, A. (2003). The role of inner speech in task switching: a dual-task investigation. *J. Mem. Lang.* 48, 148–168. doi: 10.1016/S0749-596X(02)00511-9

Feinberg, T. E., Gonzalez Rothi, L. J., and Heliman, K. M. (1986). 'Inner Speech' in Conduction Aphasia. *Arch. Neurol.* 43, 591–593. doi: 10.1001/archneur.1986.00520060053017

Gauker, C. (2011). *Words and Images: An Essay on the 'Origin of Ideas*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199599462.001.0001

Geva, S., Bennett, S., Warburton, E. A., and Patterson, K. (2011a). Discrepancy between inner and overt speech: implications for post-stroke aphasia and normal language processing. *Aphasiology* 25, 323–343. doi: 10.1080/02687038.2010.511236

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J. C., and Warburton, E. A. (2011b). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain* 134, 3071–3082. doi: 10.1093/brain/awr232

Heavey, C. L., and Hurlburt, R. T. (2008). The phenomena of inner experience. *Conscious. Cogn.* 17, 798–810. doi: 10.1016/j.concog.2007.12.006

Helm, N. (2003). *Cognitive Linguistic Quick Test*. Austin: Pro-Ed, Incorporated.

Jackendoff, R. (1996). How language helps us think. *Pragmat. Cogn.* 4, 1–34. doi: 10.1075/pc.4.1.03jac

Jones, S. R., and Fernyhough, C. (2007). Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Conscious. Cogn.* 16, 391–399. doi: 10.1016/j.concog.2005.12.003

Kertesz, A. (2006). *Western Aphasia Battery-Revised (WAB-R)*. Austin, TX: Pearson.

Kurby, C. A., Magliano, J. P., and Rapp, D. N. (2009). Those voices in your head: activation of auditory images during reading. *Cognition* 112, 457–461. doi: 10.1016/j.cognition.2009.05.007

Levine, D. N., Calvano, R., and Popovics, A. (1982). Language in the absence of inner speech. *Neuropsychologia* 20, 391–409. doi: 10.1016/0028-3932(82)90039-2

Medler, D. A., and Binder, J. R. (2005). *MCWord: An On-Line Orthographic Database of the English Language*. Available at: http://www.neuro.mcw.edu/mcword/

Morin, A. (2009). Self-awareness deficits following loss of inner speech: Dr. Jill Bolte Taylors case study? *Conscious. Cogn.* 18, 524–529. doi: 10.1016/j.concog.2008.09.008

Murray, L. L. (1999). Review Attention and aphasia: theory, research and clinical implications. *Aphasiology* 13, 91–111. doi: 10.1080/026870399402226

Oppenheim, G. M., and Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 106, 528–537. doi: 10.1016/j.cognition.2007.02.006

Oppenheim, G. M., and Dell, G. S. (2010). Motor movement matters: the flexible abstractness of inner speech. *Mem. Cogn.* 38, 1147–1160. doi: 10.3758/MC.38.8.1147

Paivio, A. (1971). *Imagery and Verbal Processes*. New York, NY: Hold, Rinehart, and Winston.

Paivio, A. (1986). *Mental representations: A Dual Coding Approach*. New York, NY: Oxford University Press.

Papafragou, A., Hulbert, J., and Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition* 108, 155–184. doi: 10.1016/j.cognition.2008.02.007

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495

Sasisekaran, J., De Nil, L. F., Smyth, R., and Johnson, C. (2006). Phonological encoding in the silent speech of persons who stutter. *J. Fluency Disord.* 31, 1–21. doi: 10.1016/j.jfludis.2005.11.005

Vygotsky, L. S. (2012/1962). *Thought and Langauge, Revised and Expanded Edition*. Cambridge, MA: MIT Press.

Watson, J. B. (1930). *Behaviorism*. Chicago: University of Chicago Press.

frontiers
in Psychology

The stream of experience when watching artistic movies. Dynamic aesthetic effects revealed by the Continuous Evaluation Procedure (CEP)

Claudia Muth [1,2]*, Marius H. Raab [1,2] and Claus-Christian Carbon [1,2]

[1] Department of General Psychology and Methodology, University of Bamberg, Bamberg, Germany, [2] Bamberg Graduate School of Affective and Cognitive Sciences, University of Bamberg, Bamberg, Germany

Research in perception and appreciation is often focused on snapshots, stills of experience. Static approaches allow for multidimensional assessment, but are unable to catch the crucial dynamics of affective and perceptual processes; for instance, aesthetic phenomena such as the "Aesthetic-Aha" (the increase in liking after the sudden detection of Gestalt), effects of expectation, or Berlyne's idea that "disorientation" with a "promise of success" elicits interest. We conducted empirical studies on indeterminate artistic movies depicting the evolution and metamorphosis of Gestalt and investigated (i) the effects of sudden perceptual insights on liking; that is, "Aesthetic Aha"-effects, (ii) the dynamics of interest before moments of insight, and (iii) the dynamics of complexity before and after moments of insight. Via the so-called Continuous Evaluation Procedure (CEP) enabling analogous evaluation in a continuous way, participants assessed the material on two aesthetic dimensions blockwise either in a gallery or a laboratory. The material's inherent dynamics were described via assessments of liking, interest, determinacy, and surprise along with a computational analysis on the variable complexity. We identified moments of insight as peaks in determinacy and surprise. Statistically significant changes in liking and interest demonstrated that: (i) insights increase liking, (ii) interest already increases 1500 ms before such moments of insight, supporting the idea that it is evoked by an expectation of understanding, and (iii) insights occur during increasing complexity. We propose a preliminary model of dynamics in liking and interest with regard to complexity and perceptual insight and discuss descriptions of participants' experiences of insight. Our results point to the importance of systematic analyses of dynamics in art perception and appreciation.

Keywords: aesthetics, Aesthetic Aha, art, dynamic appreciation, indeterminacy, ambiguity

Introduction

An Encounter with a Salami made of Wood

If you had taken a walk along a calm side road in Nuremberg last summer, you might have encountered a strange object in a window (**Figure 1**). Surprised, you might have
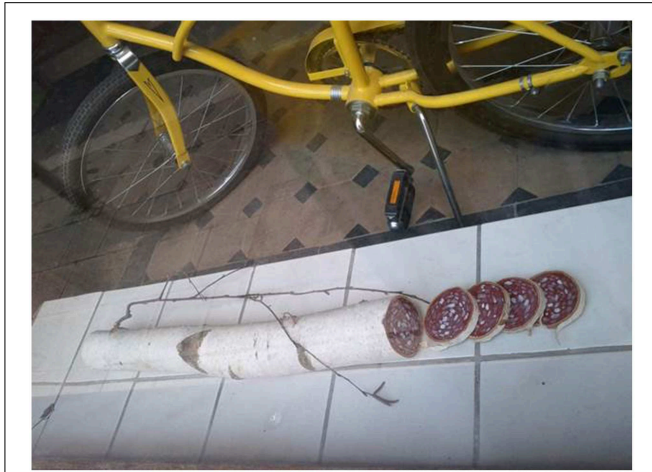
FIGURE 1 | Piece of wood, detected in a window in Nuremberg in 2013.
Photograph by Claudia Muth. Image courtesy of Claudia Muth.

stopped there wondering why one would put a big piece of sausage—actually appearing to be Milanese salami—in a sunny window. On second glance, it might have appeared to you that this object is actually not a sausage but a piece of wood cut into pieces, with the natural white of the birch bark looking like salami casing, and sausage-like patterns painted on the cut edges. This insight might furthermore have made you wonder about the function of this place: is it a gallery? And finally, despite being amused, you might have even felt a bit fooled as the illusion was obviously deliberately intended for pedestrians and made you an unwitting part of an artistic project. This little episode shows that we are guided by expectations, continuously forming predictions about the world, and that we are easily irritated when they are not met. The underlying mechanism is described in the cognitive sciences as "predictive coding"; a theory deeply rooted in the concept of perception as knowledge-driven inference proposed by psycho–physiologist Von Helmholtz (1866). Within this conceptual framework it is stated that instead of a bottom-up accumulation of information we engage constantly in matching sensory inputs with predictions created on the basis of prior experiences (for a recent critical examination of this account see Clark (2013). Artists widely make use of mismatches between predictions and actual sensory cues by providing deviations from beholders' expectations or perceptual habits. According to the *tentative prediction error account of visual art* (Van de Cruys and Wagemans, 2011) these mismatches motivate the perceiver to engage in the rewarding resolution of the prediction error.

Physical and Semantical Dynamics in Perception

Two main conclusions might follow on from this:

(a) Our perceptual impressions of an object and its context are in permanent flux as we move or as the object moves or transforms itself: the (perceived) world is not static but permanently *physically changing*.

(b) Meaning evolves out of interaction with our environment and is not obvious and inherent *per se*: the (perceived)

world is continuously re-evaluated and thus not determinate but *semantically changing*. This means that, even if something remains physically constant, our attribution of meaning is permanently updated by cognitive and affective processes. Perception is always based on psychological processing which is highly interactive and dependent of expectations, predictions, and activation of semantic networks (see Carbon, 2014).

In the realm of this research project, we examine the dynamics of such changes. These are qualities or specific patterns of changes, respectively. Dynamics can emerge out of physical changes (e.g., the typical dynamics of an explosion), semantic changes (e.g., the dynamics of a sudden shift in valence when learning about the financial value of a vase which you are holding in your hands), or their interaction (e.g., the dynamics of a perceptual Aha insight when finding Wally in a crowd after scanning the scene via eye movements). Semantical dynamics become specifically evident in the perception of objects that offer multiple meanings and afford elaboration: the "wooden sausage" (see **Figure 1**), for instance, induces dynamics specific to the phenomenon of bistable ambiguity as we switch from one determinate interpretation (sausage) to the other (wood) and eventually back again (see the psychological concept of ambiguity by Zeki, 2004, as well as the conceptualization of multistability as defined by Kubovy, 1994).

Evidently, physical and semantical dynamics are linked as we can imagine physical changes inducing semantical changes. Varini's (2006) "Huit carrés," for instance, plays with perceptual changes induced by different viewpoints of the object. This way they surprisingly reveal appearances of Gestalt from a specific physical viewpoint (see **Figure 2**). His works thus fuse physical with semantical dynamics. By concealing an identifiable pattern his work exemplifies the art theoretical definition of "hidden images" by Gamboni (2002), challenging the recipient to search (cognitively, but also by moving around or in front of the artwork) for an object that is actually present in the image but cannot be perceived too easily. A special class of semantical dynamics is found in "potential images" which—in contrast to hidden and ambiguous images—are fully indeterminate. They do not provide any recognizable object but are evocative of something we might know (for "potential images" see Gamboni, 2002; for indeterminacy see Pepperell, 2006). The example in **Figure 3** stimulates all kinds of association and motivates intense exploration without resolving into certain, determinate identification (in contrast to ambiguity, which offers several certainties with the same probability; see definition by Zeki, 2004). As the art historian Gombrich (1960) proposed for the perception of Cubist artworks—offering a wide range of indeterminacy—here also the visual search continues after cues have been detected. Indeterminacy is thus a suitable phenomenon for studying highly dynamic experiences in regard to the attribution of meaning.

Dynamics in Liking

It is quite clear that not only are dynamics to be found in the perceptual process itself and the processing of semantics, but aesthetic appreciation is obviously changeable, too. This has been an issue of empirical studies in the domain of psychological

FIGURE 2 | Varini (2006). Huit carrés. Versailles: Orangerie du Chateau de Versailles. Image courtesy of Felice Varini.

FIGURE 3 | *Succulus*, a painting made by Pepperell (2005). Image courtesy of Robert Pepperell.

aesthetics, though there remains to be a systematic examination of this process. For instance, appreciation was found to increase with unreinforced visual presentation of a stimulus (see mere-exposure effect by Zajonc, 1968) limited by the factor of increasing boredom (Bornstein, 1989) or fatigue (Carbon, 2011). While here, increasing familiarity seems to be a crucial factor, the depth of elaboration played a role in a study by Carbon and Leder (2005): innovative car designs were liked more after elaboration of the material via the so-called Repeated Evaluation Technique (RET; cf. Faerber et al., 2010). A further example of empirical evidence for dynamics relevant to appreciation is demonstrated by the phenomenon of the "Aesthetic-Aha" (Muth and Carbon, 2013), stating that the recognition of a Gestalt within two-tone-images yields a sudden increase in liking. Positive effects of the subjective strength of perceptual and cognitive insights on appreciation were also reported in the domain of art perception (Muth et al., 2015).

The mere exposure is referred to in so called fluency theories: higher familiarity by repetition leads to an increase in processing fluency which is marked by positive affect (e.g., Reber et al., 2004). The elaboration of the car interior designs—in contrast—comprised an active engagement which probably induced changes in the perception of the material's features and qualities. Here, as well as in regard to the "Aesthetic Aha" effect, dynamics in appreciation are linked to dynamics in semantics elicited by deep elaboration or increases in determinacy (Aha!), respectively. While fluency might actually also play a role in such semantical dynamics, it might be crucial that the material itself is difficult or complex enough to induce an increase in certainty in the first place for the positive effect of an "Aesthetic Aha" to occur: while in the rationale of fluency accounts we prefer the most predictable stimulus, Van de Cruys and Wagemans (2011) point to pleasure from the "transition from a state of uncertainty to a state of increased predictability" (p. 1035). The sudden revelation of a Gestalt within an indeterminate picture thus might be a rewarding resolution of indeterminacy.

Still, as the philosopher Gadamer (1960/2002) points out "There is no absolute progress and no final exhaustion of what lies in a work of art" (p. 100). Therefore, we would like to add that even if increases in predictability or certainty, respectively, are pleasurable this does not have to mean that we derive pleasure solely from arriving at a fully determinate interpretation of an artwork (if possible at all). Here, the semantical dynamics might not equal the pattern of a unidirectional progress in regard to uncertainty reduction, but instead might consist of an alternation between indeterminate phases, moments of insight, or even an endless loop among determinate patterns within ambiguous objects. Partial insights into the semantic structures of an artwork (e.g., discovering the topic of the depicted scene while still being puzzled by the choice of stylistic means) might evoke pleasure without providing "a solution" to the "problem" posed by the work (see Muth et al., 2015). Such insights might happen several times during the elaboration of an artwork and might be an important factor why many pieces of art keep the beholders' interest in them alive.

Dynamics in Interest

Aside from the actual changes in meaning attribution, the expectation of insight might also be relevant to one crucial dimension of appreciation: interest. Berlyne (1971) assumed "disorientation" with a "promise of success" to elicit interest, while Silvia (2005) proposed a combination of appraisals for interest—one being the challenging features of an object, and the other being one's ability to cope with these challenges through understanding. Thus, even before a sudden increase in determinacy, the interestingness of an object might increase. For indeterminate objects like the one depicted in **Figure 3** this might be particularly relevant: while not providing the experience of total determinacy, interest might arise due to the ongoing "promise," the permanent unresolved "potential" of determinacy.

Aims and Hypotheses

We aim at providing a more elaborate picture of the interplay of physical and semantical dynamics with dynamics of appreciation

by assessing continuous evaluations of movies in regard to the dimensions of complexity (physical dynamics), (in)determinacy and surprise (semantic dynamics), as well as liking and interest (dynamics of appreciation). We examined (i) the effects of sudden perceptual insights (increases in surprise and determinacy) on liking; that is, "Aesthetic Aha"-effects, (ii) the dynamics of interest before moments of insight, and (iii) the dynamics of complexity before and after moments of insight. Based upon the reported findings of the "Aesthetic Aha" effect (Muth and Carbon, 2013) as well as the theoretical account of reward by uncertainty reduction (Van de Cruys and Wagemans, 2011) we predict (i) an increase in liking at moments of insight, while (ii) interest might already increase before moments of insight due to the anticipation of success (Berlyne, 1971; Silvia, 2005). For a moment of insight to happen, (iii) a certain level of complexity might be necessary.

Materials and Methods

Participants

Sixty participants took part in the experiment on a voluntary basis (28 participants in a gallery setting, $mean_{age} = 38.1$ years; $range_{age} = 18–85$ years and 32 participants in a laboratory, $mean_{age} = 20.7$ years; $range_{age} = 18–24$ years). A *Snellen* eye chart test and a test with a subset of the *Ishihara* color cards assured that all of them had normal or corrected-to-normal visual acuity and normal color vision. The participants were naïve to the purpose of the study.

Apparatus and Stimuli

As material we employed "Konstrukte"—a movie (07:18 min.) by Claudia Muth (from the year 2009) which was created in an artistic context, originally not intended to be stimulus material. It depicts the evolution and metamorphosis of Gestalt (see **Figure 4** and **Supplementary Material A**) and the changes in determinacy are well suited for the study of physical dynamics (changes among subsequent film stills and their complexity) as well as semantical dynamics (emergence, disappearance, and metamorphosis of Gestalt). The artist used an intuitive drawing technique which allows for the development of Gestalt out of arbitrarily set lines to slowly reveal order in a seemingly diffuse picture and photographed the drawings (charcoal and acrylic paint) in various stages differing only slightly in detail (stop-motion technique). This way the process of drawing as well as the artist's associations can be retraced. It is suitable as material for this study because it consists of a dynamic variation of (in)determinacy; the associations of the perceiver might be forced or destroyed

and insights induced. While these preconditions might be met by static indeterminate pictures as well, utilizing a material which is dynamic in itself allows for the definition of a wide range of physical as well as semantical dynamics and their temporal comparison within and between participants' evaluations. The movie was presented on a LG W2220P screen with a 22-inch screen size and a resolution of 1680×1050 pixels.

As an input device we developed an apparatus that is able to capture assessments in a very time-accurate way. Research in visual perception science is often based on snapshots, moments, or stills of experience. Static approaches allow for differentiated and deep assessment, but are hardly able to catch the dynamics of affective and perceptual processes which—as we exemplified above—are actually crucial for certain aesthetic effects. **Figure 5** shows a broad overview on different kinds of assessment ranging from low (data measured at one time point) to high (interval of data) temporal resolution. One-time point-measurements allow for deep multidimensional assessments (e.g., the multidimensional extension of the IAT, called md-IAT, see Gattol et al., 2011). The more time points are introduced, the higher the resolution of captured dynamics; but also at the same time, the fewer the dimensions which can be included due to time constraints and problematic effects of order and fatigue. Two-time point measurements like the RET (Repeated Evaluation Technique, measuring appreciation before and after elaboration, see Carbon and Leder, 2005) can still include various dimensions but capture only coarse changes between distinct points of assessment

FIGURE 5 | Typical dimensionality and dynamics of assessments with different temporal resolution. Adapted from Raab et al. (2013).
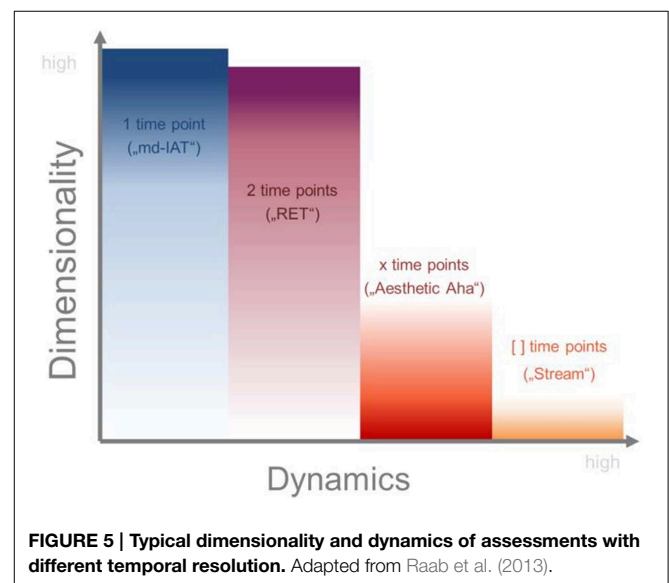
FIGURE 4 | Exemplary frames of the stop-motion movie *Konstrukte* by Claudia Muth (from the year 2009). Image courtesy of Claudia Muth.
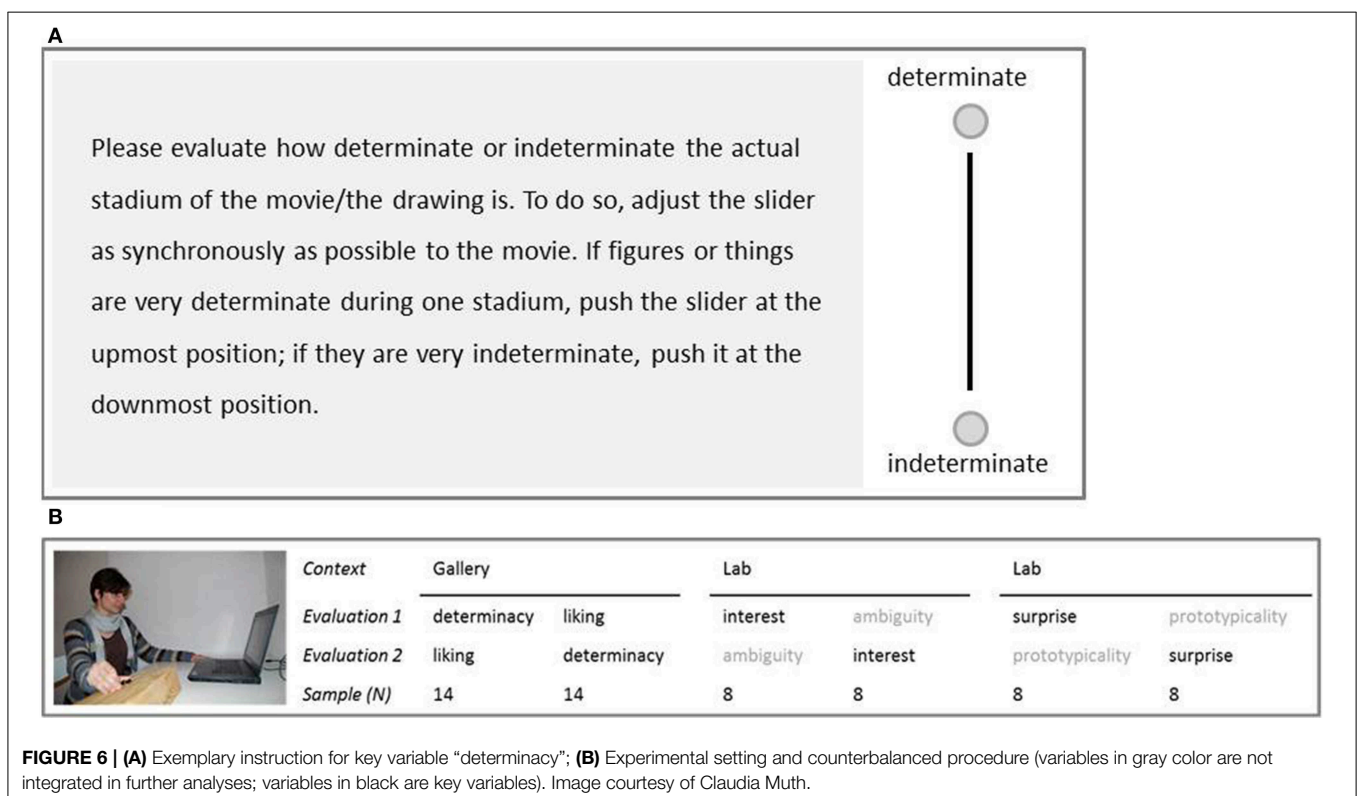
(mostly only up to 4 such time points, see Carbon et al., 2013). The "Aesthetic Aha"-effect was revealed by a finer-grained picture of changes in face clarity and changes in the liking of a picture by introducing 12 time points of measurement (Muth and Carbon, 2013). One-dimensional but temporally highly resolved assessments are achieved by, e.g., pupillometry, electro-dermal activity, or electromyography. Nevertheless, there remains a gap between such measurements and the respective qualifications of the affective states in such instances: electro-dermal activity, for instance, might be an interesting indicator of affective strength, but not of affective valence (Carbon et al., 2008).

To provide a highly dynamic assessment of evaluations of the artistic films we used a method we would like to call Continuous Evaluation Procedure (CEP; developed by the research network *Ergonomics, Psychological Æsthetics, and Gestaltung, EPÆG*), realized by employing a do-it-yourself built slider box as the input device. The CEP provides a more fine-grained picture of aesthetic assessments, and therefore also for analyzing effects like the "Aesthetic Aha" (Muth and Carbon, 2013). The system comprises a standard lever which is typically used for audio equipment (100 mm movement range, 10 kΩ linear characteristics), mounted on wooden housing. Inside the box, an ATMEGA microprocessor continuously measured the lever's resistance, mapped the resistance to a value between 80 and 1024 (with 80 indicating the lowest and 1024 the highest lever position; referred to as "strength" in the following and transposed to a scale ranging from 0 to 1 in the figures for better readability) and sent the value via an FTDI serial-to-USB converter to the connected PC. The current slider position was stored in the box as a numerical value, and updated constantly and virtually without time delay by an ATMEGA processor. Upon each new movie frame, the current value (that is, slider position) was requested via the Serial-to-USB interface. For our setup, this meant slider positions for a movie running at 30 fps could be recorded without introducing a time lag. The video presentation was realized via the Processing Library for Visual Arts and Design (Fry and Reas, 2014) and the GStreamer library (Open-Source, 2014), in which for each movie frame the current slider position was retrieved and stored. To achieve multidimensionality we used the CEP repeatedly for all key variables.

Procedure

Every participant watched the movie twice and evaluated it continuously on one dimension each time via the CEP. The instructions were given together with a graphical representation of the slider and the two poles of the according key dimension (see **Figure 6A**). Afterwards the participants were asked to push the slider up and down to get a feeling for the usability of the apparatus. One group then evaluated the key dimensions of liking and determinacy in two subsequent trials in an art gallery ("Griesbadgalerie" in Ulm, Germany; in a room separate from the exhibition, see **Supplementary Material B**). To minimize order effects, the order of the two dimensions was counterbalanced (for a visualization of the setting and the rationale of the counterbalanced design see **Figure 6B**). These dimensions were complemented by further testing sessions with other groups of people on the dimensions of surprise at an experimental laboratory at the University of Bamberg, Germany to be able to define insight moments as

FIGURE 6 | **(A)** Exemplary instruction for key variable "determinacy"; **(B)** Experimental setting and counterbalanced procedure (variables in gray color are not integrated in further analyses; variables in black are key variables). Image courtesy of Claudia Muth.

a combination of determinacy and surprise. Furthermore, we included interest as a second dimension of aesthetic appreciation by additional assessment in the same lab setting. We decided to follow this strategy as we were mainly interested in liking and determinacy and aimed to test these variables under the ecological condition of a gallery context. But to test in a gallery also means to limit the experimental approach: precisely, when testing in a gallery, the number of volunteering gallery visitors is limited, and testing runs the risk of disturbing the experience of other visitors. This made us develop the design of capturing the two key variables of aesthetic experience in the gallery and the additional variables in the laboratory (all variables were asked for in an order-balanced way). We stuck to this one-person-two-dimensions design for the gallery testing for other aesthetic factors, too, in order to keep the design consistent. After the second evaluation phase, participants filled out a questionnaire to "describe in a few own words how it felt to suddenly recognize something clearly."

Analysis

To define moments of insight, we described the movie's dynamics in four dimensions: three of them based on empirical data for determinacy, surprise, and interest along with data on complexity based on an automatized analysis (size of each frame after jpeg-compression, see Marin and Leder, 2013). We then identified moments of insight by (a) determining local maxima in the first derivative of ratings (averaged over participants) on determinacy and surprise, respectively and (b) identifying points in time where both derivatives reach common peaks. This follows the definition of insight as a sudden (strong surprise) and clear (high determinacy) solution to a problem (as proposed in the 1930s by Gestalt psychologists and redefined recently, e.g., by Bowden et al., 2005). We selected those peaks in which (a) the highest sum of both variables is achieved and (b) the sum of both variables contributes to the peak resulting in seven moments of insight (see **Figure 7A** and **Supplementary Material C**). While these peaks mark the points of biggest change in their respective dimension, we assume that the psychologically relevant event—the insight—had occurred shortly before the rapid increase detected by the CEP. By visual inspection of the yielding data curves, we determined this insight point (when the lever movement leading to the rapid change began to show in the data) to be 45 frames prior to the peak. So for any peak showing at $t = x$ (x being the movie frame), the insight was located at $t = x - 45$.

To reveal dynamics in appreciation with regard to an insight moment, we selected seven data intervals (*insight windows*) containing liking ratings ranging from 60 frames prior to each insight moment to 60 frames (which equals 4 s overall) after that insight moment and selected the according intervals of liking data (see **Figure 7**). We then phase-shifted all seven time windows around the insights to obtain one single *insight window* in which each insight moment is marked by frame "0" to be able to compare all changes in liking in relation to insight (see **Figure 8A** left). We then averaged data (see **Figure 8** middle) and used a modified cosine value of the angle between the slope describing data before (frame "−60" to frame "0") and the slope describing

data after the insight moment (frame "0" to frame "60"; see **Figure 9**).

The cosine measure is a common metric in the field of information retrieval for determining similarity between two vectors (see Singhal, 2001). It results in "1" when two (n-dimensional) vectors point in exactly the same direction; it is "0" for orthogonal vectors; and "−1" for opposing vectors. Here, we rotated and re-assembled the measure in such a way that it captures the dissimilarity between two vectors. It is "0" when both vectors (pre- and post-insight) have the same direction; it approaches "1" when the post-vector marks an increase compared to the pre-vector (where "0.5" would be an angle of 90°); and it approaches "−0.5" when the post-vector marks a decrease (**Figure 9**).

The cosine value obtained for the *insight window* thus describes changes in the corresponding variable, e.g., liking, at the insight moment. To test if this change is significantly different from the general dynamics of liking evaluations, we compared the seven cosine values at the moment of insight to those of 1000 randomly picked data intervals (*non-insight windows*) and conducted a t-test to check if the cosine values at the *insight windows* are distinguishable from the random sample.

Results

Effects of Sudden Perceptual Insight ("Aesthetic Aha")

A two-sided one sample t-test revealed that the increase in liking during an *insight window,* meaning after an insight moment, was significantly higher than other changes in liking during the evaluation of the movie, $t_{(1005)} = 2.33$, $p = 0.02$, Cohen's $d = 0.47$ (for a visual comparison to changes in adjacent windows see right panel in **Figure 8**). Changes before that time point were not significant (for a differentiated visualization of these results see **Figure 11**). Furthermore, we found a strong correlation between determinacy and liking ($r = 0.663$, $p < 0.001$; see **Figure 10**).

Dynamics of Interest before Moments of Insight

Analogous to the analysis of changes in liking in relation to insight moments, we conducted analyses on the changes in interest (with the modified cosine measure) and revealed that interest already increased 45 frames/1500 ms before the moment of insight, $t_{(1005)} = 2.53$, $p = 0.01$, Cohen's $d = 0.85$ (interestingly, even stronger than at the insight moment itself; see also **Figure 8B**, right panel and **Figure 11**). The increase was strongest 30 frames/1000 ms before the moment of insight, $t_{(1005)} = 3.78$, $p < 0.001$, Cohen's $d = 1.03$. Weaker, but still large effect sizes were found 15 frames/500 ms before the insight point $t_{(1005)} = 3.18$, $p = 0.002$, Cohen's $d = 0.75$, and at the insight moment, $t_{(1005)} = 2.64$, $p = 0.01$, Cohen's $d = 0.75$. We furthermore found a strong correlation between determinacy and interest ($r = 0.767$, $p < 0.001$).

Dynamics in Complexity before and after Moments of Insight

Plotting a phase-shifted *insight window* of complexity revealed that insights happened during an increase in complexity (**Figure 8C**).
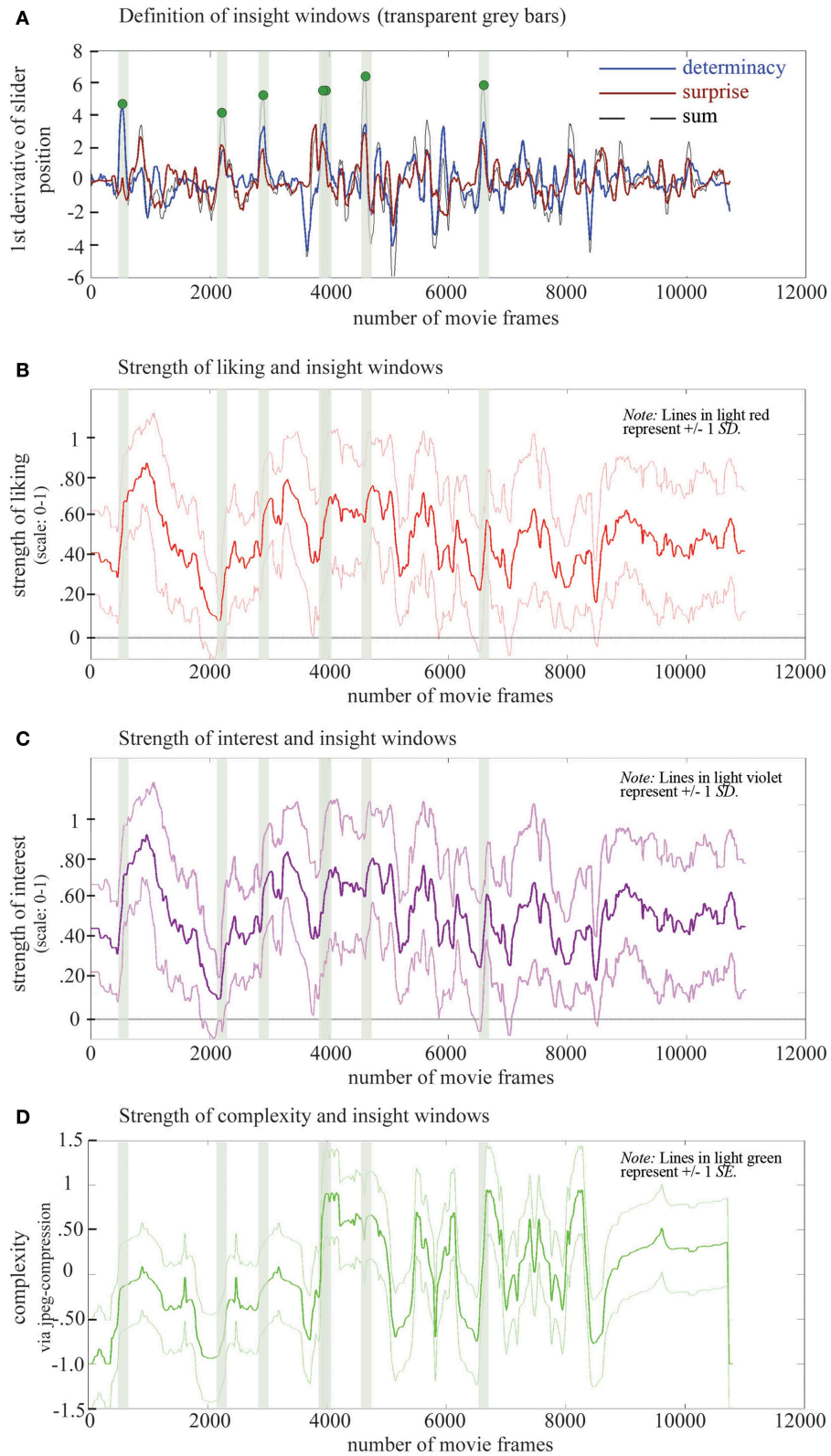
FIGURE 7 | (A) Definition of insight windows: 1st derivative of strength of determinacy and surprise (indicated by the lever position), their sum and the moments of insight, marked by green dots. **(B–D)** Dynamics in liking, interest, and complexity in relation to insight windows. Note that 2000 movie frames correspond to about 67 s.

FIGURE 8 | (A) LIKING Left panel: seven phase-shifted liking evaluations at the insight windows. Middle panel: averaged liking evaluations at the insight windows. Right panel: insight window and adjacent data windows with the only significant change being marked by an asterisk. Note: lines in light red represent the standard deviation, 2000 movie frames correspond to about 67 s. **(B)** INTEREST Left panel: seven phase-shifted interest evaluations at the insight windows. Middle panel: averaged interest evaluations at the insight windows. Right panel: insight window and adjacent data windows with the only significant change being marked by an asterisk. Note: lines in light violet represent the standard deviation, 2000 movie frames correspond to about 67 s. **(C)** COMPLEXITY Left panel: seven phase-shifted complexity evaluations (size of each frame after jpeg-compression, see Marin and Leder, 2013) at the insight windows. Middle panel: averaged complexity evaluations at the insight windows. Right panel: insight window and adjacent data windows. Note: lines in light green represent the standard error, 2000 movie frames correspond to about 67 s.

FIGURE 9 | The modified cosine theta measure (Θ) to capture dissimilarity between pre- and post-insight. The data section before the insight as well as after the insight was approximated with a line each (top left). Between these two vectors, the angle was determined. Two vectors with exactly the same direction (i.e., with no difference in direction pre- and post-insight) result in an angle of 0° between vectors and thus in a cosine theta measure of 0. A directional change upwards (second vector pointing higher than the first one) will result in a positive theta (e.g., 0.5 for a 90° upward angle between vectors); a directional change downward in a negative theta value (e.g., −0.5 for a 90° downward angle).

FIGURE 10 | Strength of determinacy and liking. The difference of both variables is indicated by gray colored areas.

Discussion

We examined (i) effects of sudden perceptual insights (increases in surprise and determinacy) on liking; that is, "Aesthetic Aha"-effects, (ii) dynamics of interest before moments of insight, and (iii) dynamics of complexity before and after moments of insight while watching an artistic movie. Our findings showed that the analysis of a continuous stream of data can reveal dynamic relationships between an artwork's physical and semantical dynamics and appreciation. Such analyses demonstrate that aesthetic experiences unfold in a dynamic way: (i) insights indeed elicited

an "Aesthetic Aha," an increase in liking (as in Muth and Carbon, 2013). (ii) The rise in interest before an insight supports Berlyne's (1971) as well as Silvia's (2005) ideas that interest is evoked by (the appraisal of) challenge and coping potential—the expectation of success. (iii) Furthermore, it seems that for an insight to occur, stimuli have to possess a certain complexity. In dynamic terms this means that only phases in the movie in which complexity increases have the potential to lead to insightful moments—which again evoke an increase in liking. This is in accordance with the idea proposed by Van de Cruys and Wagemans (2011) that it is not predictability itself

but the reduction of uncertainty induced by prediction-errors that might bring pleasure. There are two directions of interpretation of these three results: either (a) the Aha-insight benefits from, or is more probable because of, an orienting reaction and high interest due to an increase in complexity or (b) interest arises due to "affective forecasting" ("people's predictions about their future feelings," see (Wilson and Gilbert, 2003); Wilson and Gilbert, p. 346) of this Aha-insight. In addition, the experience of an Aha-insight itself might stimulate the orienting reaction and deeper elaboration of subsequent phases as is exemplified by descriptions of participants, for instance the following: "proud of myself, motivated to continue watching (maybe one

FIGURE 11 | *Insight window* **revealing the dynamics of liking and interest in relation to moments of insight.** An exemplary angle from which we derived the cosine value that we compared to a random set of 1000 cosine values of other changes is depicted for liking at the moment of insight. Numbers signify the strength of the effect via Cohen's d for significant changes at the insight moment [Δ(frames) = 0] as well as prior to the moments of insight [at Δ(frames) = −15, Δ(frames) = −30, and Δ(frames) = −45]. Note: changes which are significant at an α-level of 5% are marked by an asterisk.

detects something else/more)," "I concentrated on the monitor and every recognition of a "picture" pleased me and incited me to search even more "intensely" (with more concentration) for more "pictures."

Such links between perception, affection, and appreciation suggest that systematic analyses of dynamics in art perception are crucial for an understanding of the unfolding of meaning as well as the experience of art in general. Still, people's descriptions of their experiences of an Aha-insight show that there might be additional mechanisms involved in the links between complexity, insight, and appreciation. To stimulate further discussion, we exemplify three noticeable aspects which were frequently present in the *post-hoc* descriptions of the participants' own experiences of Aha-insight moments in **Table 1**. One aspect regards descriptions of familiarity or related concepts such as "control," "relatedness," "success," or "relieve" which might point to the appraisal of coping potential (see 1st column in **Table 1**). Another less frequently mentioned aspect is the relationship between expectation and resulting recognition, and their effect on appreciation. Two different possibilities were present in the descriptions: prediction confirmation as well as prediction error were experienced as pleasurable (see 2nd column in **Table 1**). A further aspect regards the distinction between process-focused and result-focused elaboration, when participants mentioned "transitions" between Gestalt perception or reported having enjoyed the process itself vs. when they described the search for recognizable Gestalt (see 3rd column in **Table 1**).

The relationship between the prediction of a Gestalt and actual recognition of a Gestalt seems to have a very different effect on appreciation for different individuals: whereas some participants described a confirmation to be pleasurable, others clearly preferred surprising transitions within a movie. This difference might be very important in the domain of art perception as it seems to imply that the valence of any experience of new insights is mediated by perceptual and cognitive habits.
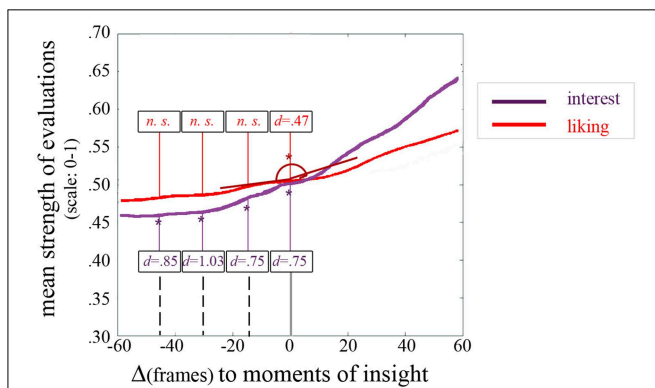
TABLE 1 | Examples of participants' descriptions of their experiences of Aha Insight moments, translated by the authors.

Coping	Prediction confirmation vs. prediction error	Process-focus vs. result-focus
"A feeling of 'familiarity'"	"[…] can be negative then, if the picture (where one recognizes something) does not conform to the expectation."	"Pleasing and partially surprising. The development into a new form, the process of transition is however more interesting than the final result (in this case the clearness)."
"Insight, relief, higher relatedness to the picture"	"Pleasing because assumption got confirmed, then a little bit disappointed that it was what one suspected."	"[…] also it is fascinating to recognize something new, especially if you consider what it evolved out of."
"Relieved and maybe a bit proud"	"[…] at the same time also happy or less happy about what I saw. Sometimes also disappointed, that something other than what I thought would come became recognizable."	"It was comfortable to detect things and to see them evolving."
"Certain experience of success, positive feeling"	"[…] with the structure getting clearer and the result more foreseeable it got more 'boring'."	"I was excited by the change."
"Additionally, the recognizable contours gave me a feeling of security or also 'control'."	"Partially it is also frustrating, if you have waited for the drawing to develop but could not recognize or classify anything."	"It was interesting to recognize something out of 'nothing.' To identify something suddenly felt clearly better than the transitions among the different motives."

All of these factors are potentially different in gallery visitors compared with volunteers, mostly art-naïve students, typically tested in laboratory conditions. Another speculation concerns the mode of perception: we might focus on the process of physical and semantical transformations or on the resulting recognizable Gestalt, respectively. It would be highly interesting to systematically and explicitly investigate in future studies how the activation of these modes of processing is related to personality and context factors.

The assessment of dynamics in perception provides insights not only in regard to continuous changes and dynamic relationships but also enables us to reveal effects of expectation. We compared the two groups of participants evaluating the surprise induced by the different phases in the movie either during its first or its second presentation. The pattern of the resulting data implied that it makes a difference if an observer expects the development of Gestalt or indeterminacy due to previous exposure or if predictions are formed by the available information only. As these differences neither seem to be systematic nor easily describable, e.g., by a shift of peaks to earlier phases, this point is left for following studies with a concise focus on effects of expectation.

Art is not only an ideal medium through which physical and semantical dynamics in experience can be studied: while artworks might be insightful for perception science as they do not represent things as they are but as they are perceived (Fiedler, 1971), many artworks also do explicitly refer to and reflect on these dynamics. Some of Paul Cézanne's works hinder determinate interpretation and instead provide potential shapes of things. Gamboni (2002, p. 116) thus states that Cézanne leaves images "in perceptual formation and makes the spectator conscious of the interpretative process in which he is engaged and which will never be conclusive." Also Majetschak (2003) related Cézanne's works to the constructive activity of perception and called them consequently a "birthplace of visibility" (translated from "Geburtsort von Sichtbarkeit," p. 324). The dynamics of potential images (Gamboni, 2002) are also found in Cubist artworks by, e.g., Picasso and Braque which never provide full determinacy but contradictory cues which evoke an ongoing search for Gestalt (Gombrich, 1960). Furthermore, a Cubist artwork might allow for the retracing of part of the artist's dynamic elaboration of the original object: instead of a fixed spatial relation between painter and object, Cubist artists applied a "mobile perspective" (e.g., Metzinger, 1910), an "analytic description" of the object, or a "synthesis" of various viewpoints (translated from Kahnweiler, 1971, p. 69). This simultaneity of spatial dynamics within one picture might let us simulate an actual visual exploration which integrates several fragmented "semi-worlds" (Churchland et al., 1994)—inhomogeneous of detail and colorization—over time. Cubist artworks might thus be good examples of two kinds of semantical dynamics induced by the simultaneity of both potential identifiable forms and perspectives. Examples of the capturing of dynamics by artworks often regard the illusory layer of the artwork (the depicted objects or sceneries)—not the material layer (canvas and color). But semantical dynamics can also evolve in regard to the material itself. The carpets made by Faig Ahmed (**Figure 12**) are valuable examples of semantical

dynamics due to switches between interpretations of the material (paint vs. carpet) along with other semantical levels like the historical, social, and cultural dimensions of traditional carpet weaving. Contemporary "Relational art" (Bourriaud, 2002) expands the dynamic relationship between artwork and perceiver even more as it points to or includes the social context of behavioral interactions. The Munich based group *Die Urbanauten*, for instance, organizes "Swarm-Happenings" in which people are instructed to fulfill urban interventions—actions in public space like having a picnic on a bridge—to reflect on and inspire discussions about how public space is used and the set of norms which underlie our behavioral variety. Works of art thus neither have to be exhibited in a well-defined artistic context (gallery, museum) and contemplated by one observer alone, nor is the authorship and evaluation of these works independent of the social context. This viewpoint opens up a variety of possible dynamic levels that art perception comprises; the presented investigation is a first step into this wide and thrilling field of art and research.
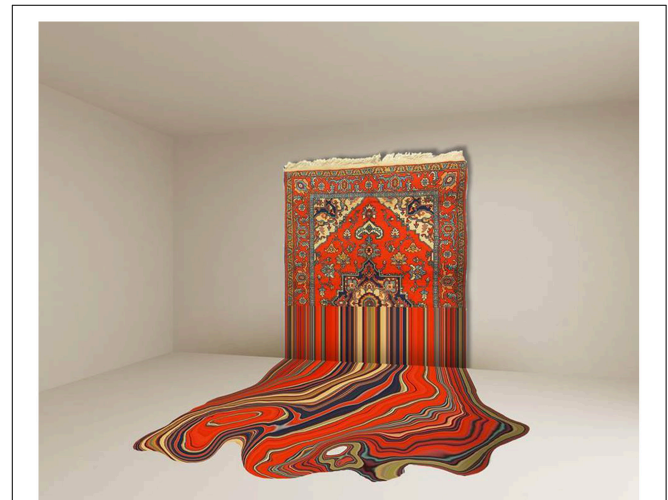

FIGURE 12 | *Liquid* **by Faig Ahmed in the year 2014 [Handwoven rug made of wool].** Image courtesy of Cuadro Gallery, United Arab Emirates.
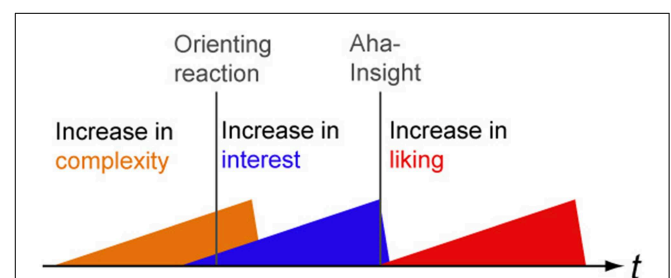

FIGURE 13 | A preliminary model of physical and semantical dynamics in interest and liking. An increase in complexity elicits an orienting reaction along with an increase in interest which precedes the Aha Insight ("Aesthetic Aha," Muth and Carbon, 2013).

Outlook: A Preliminary Model for Physical and Semantical Dynamics in Liking and Interest

Based on our findings we would like to postulate a preliminary model of physical and semantical dynamics in the perception of indeterminate material and its effects on liking and interest (see **Figure 13**): an increase in complexity might signal the potential meaningfulness of a situation or object which makes this case more interesting. Such a tag on interest triggers an orienting reaction which unleashes further cognitive resources to process the potentially relevant item. By at least partly solving indeterminacy, a Gestalt is recognized which allows an Aha-insight to occur. The result of such an "Aesthetic Aha" is an increase in liking (as in Muth and Carbon, 2013).

Conclusion

Perception and appreciation are evidently dynamic—and thus should be investigated by means of measures capturing such dynamics. This not only holds for dynamic stimulus material but also for the perception of a static object as it includes physical changes (for instance adapting one's own posture) and potential changes in semantics and appreciation over time and elaboration. Our results point to the importance of systematic analyses of such dynamics in art perception and appreciation by grasping the continuous nature of such experiences. We hope our preliminary model of dynamics in liking and interest, and their relation to complexity and perceptual insight inspires further research on the dynamics of perception and appreciation.

Acknowledgments

We would like to thank Lisa Aufleger, Stefan Breitschaft, Janina Plessow, and Johanna Günzl for assistance in testing and Alun Brown for proofreading the text.

Ethical Statement

Before the experiment, participants gave written consent for participating in the study. After the experiment had ended participants were fully informed about the aims of the study and had the opportunity to ask questions. All data were collected anonymously and no harming procedures were used. Ethical approval of the study was provided by the local ethics committee ("Ethikrat der Otto-Friedrich-Universität Bamberg"; dated 13 March, 2015).

Supplementary Material

Supplementary Material A | The movie utilized as stimulus material can be watched here: http://vimeo.com/46138003.

Supplementary Material B | Impressions regarding the exhibition at the "Griesbadgalerie" can be gained here: http://vimeo.com/68991518.

Supplementary Material C | A visualization of the development of moments of insight can be found here: https://janus.allgpsych.uni -bamberg.de/CEP_Revision/insightlow.html.

References

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York, NY: Appleton-Century-Crofts.

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychol. Bull.* 106, 265–289. doi: 10.1037/0033-2909.106.2.265

Bourriaud, N. (2002). *Relational Aesthetics*. Dijon: Les presses du réel.

Bowden, E. M., Jung-Beeman, M., Fleck, J., and Kounios, J. (2005). New approaches to demystifying insight. *Trends Cogn. Sci.* 9, 322–328. doi: 10.1016/j.tics.2005.05.012

Carbon, C. C. (2011). Cognitive mechanisms for explaining dynamics of aesthetic appreciation. *i-Perception* 2, 708–719. doi: 10.1068/i0463aap

Carbon, C. C. (2014). Understanding human perception by human-made illusions. *Front. Hum. Neurosci.* 8:566. doi: 10.3389/fnhum.2014.00566

Carbon, C. C., Faerber, S. J., Gerger, G., Forster, M., and Leder, H. (2013). Innovation is appreciated when we feel safe: on the situational dependence of the appreciation of innovation. *Int. J. Des.* 7, 43–51.

Carbon, C. C., and Leder, H. (2005). The repeated evaluation technique (RET): a method to capture dynamic effects of innovativeness and attractiveness. *Appl. Cogn. Psychol.* 19, 587–601. doi: 10.1002/acp.1098

Carbon, C. C., Michael, L., and Leder, H. (2008). Innovative concepts of car interiors measured by electro-dermal activity (EDA). *Res. Eng. Des.* 19, 143–149. doi: 10.1007/s00163-008-0045-2

Churchland, P., Ramachandran, V. S., and Sejnowski, T. (1994). "A critique of pure vision," in *Large-Scale Neuronal Theories of the Brain,* Eds C. Koch and J. Davis (Cambridge, MA: MIT Press), 23–66.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Faerber, S. J., Leder, H., Gerger, G., and Carbon, C. C. (2010). Priming semantic concepts affects the dynamics of aesthetic appreciation. *Acta Psychol.* 135, 191–200. doi: 10.1016/j.actpsy.2010.06.006

Fiedler, K. (1971). *Schriften zur Kunst*. Munich: Fink.

Fry, B., and Reas, C. (2014). *Processing Library for Visual Arts and Design*. Available online at: http://www.processing.org

Gadamer, H.-G. (1960/2002). *Truth and Method*. New York, NY: Continuum.

Gamboni, D. (2002). *Potential Images: Ambiguity and Indeterminacy in Modern Art*. London: Reaktion Books.

Gattol, V., Sääksjärvi, M. C., and Carbon, C. C. (2011). Extending the implicit association test (IAT): assessing consumer attitudes based on multi-dimensional implicit associations. *PLoS ONE* 6:e15849. doi: 10.1371/journal.pone.0015849

Gombrich, E. (1960). *Art & Illusion: A Study in the Psychology of Pictorial Representation*. London: Phaidon.

Kahnweiler, D.-H. (1971). *Der Gegenstand der Ästhetik*. München: Heinz Moos Verlag.

Kubovy, M. (1994). The perceptual organization of dot lattices. *Psychon. Bull. Rev.* 1, 182–190. doi: 10.3758/BF03200772

Majetschak, S. (2003). "Die modernisierung des blicks. Über ein sehtheoretisches motiv am anfang der modernen kunst," in *Die Kunst der Wahrnehmung. Beiträge zu einer Philosophie der Sinnlichen Erkenntnis,* ed M. Hauskeller (Kusterdingen: Die graue Edition), 298–349.

Marin, M. M., and Leder, H. (2013). Examining complexity across domains: relating subjective and objective measures of affective environmental scenes, paintings and music. *PLoS ONE* 8:e72412. doi: 10.1371/journal.pone.0072412

Metzinger, J. (1910). "Note on painting," in *Art in Theory. 1900–2000. An Anthology of Changing Ideas,* eds C. Harrison and P. Wood (Oxford, UK: Blackwell), 184–185.

Muth, C., and Carbon, C. C. (2013). The aesthetic aha: on the pleasure of having insights into Gestalt. *Acta Psychol.* 144, 25–30. doi: 10.1016/j.actpsy.2013.05.001

Muth, C., Hesslinger, V., and Carbon, C. C. (2015). The appeal of challenge in the perception of art: how ambiguity, solvability of ambiguity and the opportunity for insight affect appreciation. *Psychol. Aesthetics, Creativity, and the Arts.* 9.

Open-Source. (2014). *GStreamer 1.4.2.* Available online at: http://gstreamer.freedesktop.org/

Pepperell, R. (2006). Seeing without objects: visual indeterminacy and art. *Leonardo* 39, 394–400. doi: 10.1162/leon.2006.39.5.394

Raab, M. H., Muth, C., and Carbon, C. C. (2013). "M5oX: methoden zur multidimensionalen und dynamischen erfassung des nutzererlebens," in *Mensch & Computer 2013 – Workshopband,* eds S. Boll, S. Maaß, and R. Malaka (Munich: Oldenbourg), 155–163. doi: 10.1524/9783486781236.155

Reber, R., Schwarz, N., and Winkielman, P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Pers. Soc. Psychol. Rev.* 8, 364–382. doi: 10.1207/s15327957pspr0804_3

Silvia, P. J. (2005). Cognitive appraisals and interest in visual art: exploring an appraisal theory of aesthetic emotions. *Empir. Stud. Arts* 23, 119–133. doi: 10.2190/12AV-AH2P-MCEH-289E

Singhal, A. (2001). Modern information retrieval: a brief overview. *IEEE Data Eng. Bull.* 24, 35–43.

Van de Cruys, S., and Wagemans, J. (2011). Putting reward in art: a tentative prediction error account of visual art. *i-Perception* 2, 1035–1062. doi: 10.1068/i0466aap

Von Helmholtz, H. (1866). *Handbuch der Physiologischen Optik: Mit 213 in den Text Eingedruckten Holzschnitten und 11 Tafeln.* Leipzig: Voss.

Wilson, T. D., and Gilbert, D. T. (2003). Affective forecasting. *Adv. Exp. Soc. Psychol.* 35, 345–411. doi: 10.1016/S0065-2601(03)01006-2

Zajonc, R. B. (1968). Attitudinal effects of mere-exposure. *J. Pers. Soc. Psychol.* 9, 1–27. doi: 10.1037/h0025848

Zeki, S. (2004). The neurology of ambiguity. *Conscious. Cogn.* 13, 173–196. doi: 10.1016/j.concog.2003.10.003

Emotionally excited eyeblink-rate variability predicts an experience of transportation into the narrative world

Ryota Nomura[1]*, Kojun Hino[2], Makoto Shimazu[3], Yingzong Liang[4] and Takeshi Okada[1]

[1] *Faculty of Education, The University of Tokyo, Tokyo, Japan,* [2] *College of Arts and Sciences, The University of Tokyo, Tokyo, Japan,* [3] *Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan,* [4] *Graduate School of Engineering, The University of Tokyo, Tokyo, Japan*

Collective spectator communications such as oral presentations, movies, and storytelling performances are ubiquitous in human culture. This study investigated the effects of past viewing experiences and differences in expressive performance on an audience's transportive experience into a created world of a storytelling performance. In the experiment, 60 participants (mean age = 34.12 years, SD = 13.18 years, range 18–63 years) were assigned to watch one of two videotaped performances that were played (1) in an orthodox way for frequent viewers and (2) in a modified way aimed at easier comprehension for first-time viewers. Eyeblink synchronization among participants was quantified by employing distance-based measurements of spike trains, D^{spike} and $D^{interval}$ (Victor and Purpura, 1997). The results indicated that even non-familiar participants' eyeblinks were synchronized as the story progressed and that the effect of the viewing experience on transportation was weak. Rather, the results of a multiple regression analysis demonstrated that the degrees of transportation could be predicted by a retrospectively reported humor experience and higher real-time variability (i.e., logarithmic transformed SD) of inter blink intervals during a performance viewing. The results are discussed from the viewpoint in which the extent of eyeblink synchronization and eyeblink-rate variability acts as an index of the inner experience of audience members.

Keywords: eyeblink-rate variability, eyeblink synchronization, transportation, viewing experience, Rakugo, expert

Introduction

Collective spectator communications such as oral presentations, movies, and storytelling performances are ubiquitous in human culture. Spectators who share time and space frequently involve their minds and bodies in fascinating performances. Some spectators would describe their experience as being 'carried away' by the story. This engrossing temporal experience is known as "transportation into the narrative world" (Sestir and Green, 2010). In a previous study, researchers summarized facilitators of narrative transportation (Van Laer et al., 2014). For instance, Van Laer et al. (2014, p. 803) and pointed out that stories containing more identifiable characters to audience members, plotlines that storytelling audiences can imagine, and verisimilitude all increase the likelihood that a narrative transportation will occur. In addition, an audience

member's familiarity with a story topic, attention level, transportability (i.e., "a story receiver's chronic propensity to be transported," see Van Laer et al., 2014, p. 804), age, education, and gender (female rather than male) all play a role in the likelihood of narrative transportation. Though these studies have focused on human traits—in other words, static factors of transportation—dynamic factors such as fluctuation between attention allocation and attention release during a performance also affect a transportive experience in live theater, as expressiveness between a performer and the audience is communicated in real-time. However, the processing mechanism by which an audience experiences transportation through the appreciation of expert performances remains a mystery.

Investigations into audiences' transportive experience during a storytelling performance have suggested that audience attention tends to synchronize with the addition of audio-visual stimuli used during expert performances (Nomura and Okada, 2014). Nomura and Okada (2014) showed that during an expert performance, eyeblinks among participant audience members synchronized with greater frequency and more intensity compared to audience members of a novice performance, even though the expert and novice performers performed the same story. At the same time, subjective rating scores on a scale to determine transportation into the world of the story (Miall and Kuiken, 1995) including somatic responses (e.g., sweat and chills, Nomura, 2013) were much higher for participants who watched an expert performance than those who watched a novice performance. The timing of eyeblinks is interrelated within attentional process (Nakano et al., 2009; Nomura and Okada, 2014). In general, people search for a target upon which to focus their attention. If audiences find a target, they allocate their attention to it. After this focused concentration, they release their attention to prepare to search for the next target. Audience eyeblinks decrease at the moment of attention allocation while they increase at the moment of attention release. Therefore, eyeblinks tend to synchronously occur at implicit attentional breakpoints among readers while reading books (Hall, 1945) and among viewers while viewing videos (Nakano and Kitazawa, 2010). An additional qualitative analysis (Nomura and Okada, 2014, Study 2) also indicated that eyeblinks among audiences are synchronized corresponding to scene changes and high points of expressive performance. This externally coordinated attention leads to an efficient cognitive process by avoiding loss of significant information (Nakano et al., 2009). Thus, the authors concluded that eyeblink synchronization among audiences is guided by an expert performance created to make audiences comprehend the important information.

However, it is unclear how eyeblink synchronization among audiences relates to their experience of transportation. One of the possible mechanisms is that eyeblink synchronization among audiences is driven by attentional cycles, which are in turn driven by emotional processing. One's eyeblinks usually cycle in self-paced (physiological) periods with some fluctuations. However, audience eyeblink onset might be delayed or accelerated depending on the actors' expressions, as the timing of attention allocation and attention release are coordinated with the performance. When audience members shift their attention back and forth more frequently, in parallel

with the storyline and punchlines performed by the actors, eyeblink time points vary dynamically, but sensitively, in line with the performance (Nomura and Okada, 2014). As s result, eyeblinks among audiences synchronize with each other. Because the duration of attentional cycles reflects the audience's active involvement in a performance, durations vary more frequently than those of self-paced cycles. Such dynamic attention shifts bring audience members more emotional excitement. Their emotional excitement may motivate them to pay attention for upcoming expressions that could contain important content-related information. Thus, a reciprocal process between emotional excitement and the resulting motivated attention would affect transportative experiences. In other words, high emotional excitement and high eyeblink variability would predict that audience members experience more transportation.

The other possibility is that a situation model improves prediction accuracy and simultaneously facilitates the experience of transportation. A situation model refers to a reader's representation of the referents and events described in a text (Van Dijk and Kintsch, 1983). More generally, it refers to a story receiver's mental model (Johnson-Laird, 1983) using specific information that aids the comprehension of the current situation. When people comprehend a story, they construct a representation of the situation and its words and sentences (Zwaan et al., 1995). The current situation model manages new information from the aspect of temporal, spatial, or casual consistency (Radvansky and Copeland, 2001) and possibly enables people to predict the next plot twist more precisely. If an audience can construct representations of a story, they will more easily understand the meaning of the situation. In other words, a situation model reduces the cognitive burden required to comprehend a story. At the same time, this reduction facilitates the experience of transportation, because audiences can freely use remaining cognitive resources for other cognitive activities, such as focusing on the detail of expressions. Thus, a model can help an audience realize the depth of feelings expressed in a performance. While a non-experienced audience constructs a situation model by using only the knowledge accumulated through appreciation of the present story, an experienced audience constructs a model by also exploiting domain knowledge cultivated through past viewing experience. In light of this perspective, it could be predicted that the experienced audience, compared to the non-experienced audience, would gain more transportative experience from the beginning of a performance.

In summary, the mechanisms of audiences' eyeblink synchronization reflecting the experience of transportation are as follows. On one hand, externally coordinated attention leads to dynamic eyeblink shifting, as well as emotional processing, due to which audience members are inclined to pay additional attention to the performance. On the other hand, a mental model reduces the cognitive burden of comprehending characters and plotlines of a story, while simultaneously improving the accuracy of prediction. These two mechanisms facilitate audience eyeblink synchronization. However, these mechanisms could be interdependent. In general, synchronizations caused by external inputs are possible if respective elements respond reliably to time-varying

stimuli (Mainen and Sejnowski, 1995). Thus, eyeblink synchronization among audiences during Rakugo (traditional Japanese vaudeville) settings could occur owing to performance quality in addition to audience sensitivity to external stimulus. For instance, even though emotional excitement and biased distribution of eyeblinks predict the transportive experience, this result may be obtained from the experienced audiences only if domain knowledge is a necessary condition. In another case, even the non-experienced audiences may obtain a transportive experience if the performance contains sufficient information to guide their attentional process. The purpose of this study is to examine these two potential mechanisms and their relationship to each other. In the experiment, experienced and non-experienced audiences were assigned to watch one of two videos separately: an orthodox performance (played in front of frequent viewers) or a modified performance (played in front of first-time viewers) acted by the same artist (The details of the two performances will be described later). In all settings, participants' eyeblink responses were observed.

The time cycle of inter-blink intervals (IBIs) varies when the performance contains more frequent expressions that draw the viewer's attention, because the original (self-paced) period becomes accelerated or delayed. This leads to a higher rate of eyeblink variability. Thus, the SD of IBIs can be used as a measurement of eyeblink-rate variability on an individual level. Furthermore, emotional excitement can be measured by a subjective rating score on a humor scale, while it is no simple task to measure each audience member's situation model *per se* during real-time processing. However, the similarity of situation models among audiences could be estimated by focusing on the reproducibility of participants' eyeblink responses, because eyeblinks by audiences who have a common situation model would unintentionally select similar information, leading to more closely timed (i.e., more reliable) and more similar eyeblink patterns. In this study, we observed the precision of eyeblink responses by focusing on time differences within an audience, instead of defining the objective criteria or identifying audio-visual information to which an audience allocates its attention. We calculated mean eyeblink timing asynchrony and estimated mean similarity of IBI patterns between two particular audience members as group-level indices of reproducibility.

To investigate the first mechanism, we performed a multiple regression analysis with a SD of IBIs and humor ratings as an explanatory variable with a self-reported degree of transportation as a target variable. This hypothetical mechanism is rejected when SDs of IBIs or a subjective-humor response have no predictive power. Here, we were unable to eliminate the possibility that other variables suggested by previous research (Van Laer et al., 2014) were also facilitating or inhibiting the process interdependently. In the multiple regression analysis, we included age, gender, mean of IBIs calculated for each individual, knowledge of the performer (a dummy variable), and knowledge of the story (a dummy variable) as possible predicting variables. This analysis was performed across the experimental conditions, with the aim of determining whether variables of real-time

processing, rather than other variables such as the nature of performance or the different viewing experiences, had a predictive power for transportive experience. As the first hypothetical mechanism was supported by multiple regression analysis, we went on to examine the second mechanism, concerning the use of a situation model. If the asynchrony of eyeblinks was lower in the experienced audience group than in the non-experienced group, it would suggest that domain knowledge had helped in the construction of a situation model. One-way ANOVA was performed to assess the interaction between viewing experience and actor expression during performance. If a situation model was unnecessary for transportation, the degree of transportation did not increase—at least in the situation in which group-level eyeblink asynchrony was high. If any other factor was suspected of contributing to the process of transportation throughout the analyses, an additional analysis was performed according to the nature of stimuli such as laughter of audience recorded during Rakugo (Japanese) vaudeville performances (i.e., *in situ*).

Materials and Methods

Participants

Participants included 28 males and 44 females, all native Japanese speakers. Out of 72 people who participated in the experiments, complete eye-tracking data was obtained from 60 participants (24 males and 36 females, mean age = 34.12 years, range = 18–63 years). Eye-movement data from two people was not usable due to drooping eyelids. Data from 10 other people were unusable because troubles with the instruments caused a loss of eye-detecting information leading to insufficient records. The experimenter defined participants who had viewed the type of storytelling performance used in the study more than 10 times in any situation, including through other media and live performances, as an experienced audience member. The experimenter adopted the criteria because the mean number of viewing times was usually three or four times in the daily lives of most Japanese. This meant that individuals who met the criteria seemed to seek opportunities to view Rakugo more than five times. As a result, 30 (15 experienced and 15 non-experienced) participants were assigned the videotaped performance as first-time viewers and the remaining 30 participants (15 experienced and 15 non-experienced) were assigned the videotaped performance as frequent viewers.

The Storytelling Artist and Stimuli

In the current study, the authors asked professional Rakugo artist Kokontei Bungiku (34-year-old performer with 10 years' experience) to record his performances. Rakugo is a traditional Japanese comic vaudeville storytelling performance in which one artist plays many characters. The stage setting is usually just a square cushion (*zabuton*) on which the performer sits to tell passed-down and newly created stories. The artist uses a Japanese fan and a traditional hand towel to represent all stage properties such as chopsticks and a sword (*katana*). In a traditional Rakugo apprenticeship of the Association of Rakugo (General Incorporated

Associations), the title of first-rank performer ("Shin'uchi") was given to Bungiku earlier than the 28 senior performers. We therefore assumed that Bungiku possesses the skills to modify his performance according to the nature of the audience. Two storytelling performances as well as the audio-visual information during the performance were videotaped. In both performances, the story Bungiku told was called "Nibansenji," literally meaning the second brew of tea or decoction, which is semantically transferred to mean that things become a pale imitation. The outline of the story is as follows: five civilians go around the city of Edo (the old name of Tokyo) to prevent fires on a very cold winter night. After enduring the cold, they go back to a hut and have a warm meal, while passing around a small cup of warmed sake, conveniently concealed as "senji-kusuri" (decoction). Suddenly, a samurai who supervises the fire-prevention activities comes to the hut and calls for the door to be opened. Although civilians hurriedly hide the meal and sake, the samurai notices them quickly and wants to make them his own, relying on his authority. While the samurai wants another decoction (i.e., sake), one of the civilians answers that they have no more decoction. As the last line, the samurai orders as follows: "While I patrol around the neighborhood, brew the second decoction."

The performances were recorded on December 6, 2013, in a Rakugo vaudeville setting that was recreated in a laboratory room at the University of Tokyo. The artist performed live in front of 31 frequent viewers and 24 first-time viewers. They (i.e., audience *in situ*) were different from participants of the current laboratory experiment. Several experimenters and assistants were also present in the room. The performance for frequent viewers was acted in the style of traditional vaudeville storytelling performance in everyday theater (orthodox video). The performance for first-time viewers was played in a modified way to help first-time viewers better comprehend the content of the story (modified video). The videos lasted approximately 3220 s (53 min 40 s) and 3022 s (50 min 22 s), respectively. For the first-time viewers, the artist took a few minutes to explain the traditional way of viewing this type of storytelling performance.

The videotaped performance was presented on a 19-inch monitor distanced 58 cm from each participant. The video was projected to a size of 15 cm (H) × 24.6 cm (W). The subjects' viewing angle of the performer, who was sitting on the *zabuton*, was approximately 11.3° × 10.7° located at the center of the monitor. The projected size was approximately equivalent to the size of performers viewed by an audience seated at a 5-meter distance in the center of a vaudeville theater. The video stimuli was controlled by a desktop personal computer (Dell, Optiplex 900, CPU 3.40 Ghz, Memory 8.00 GB).

Eye movements were measured by a non-contact, eye-tracking device (EMR-AT, VOXER, nac Image Technology Inc.) at a sampling rate of 60 Hz. The eye position was smoothed using a moving-average method and recorded electronically. The eyeblinks were detected by instantaneous losses (0.3–1.0 s) of pupil with an eye-position motion that went rapidly down and then immediately up. The first time point during the detected eyeblink was identified as the onset of that eyeblink. The time duration from one onset to the next onset was defined as the IBI. Each

participant's chin and forehead were placed in fixed way on a support device to minimize the influence of head movements on eye-tracking data. Presentation of stimuli were controlled and recorded by a background program made by Visual Basic. A few time delays occurred before the presentation while the computer was loading a video. These presentation time delays were corrected using recorded time stamps.

Questionnaire

The questionnaire package consisted of two scales (humor and transportation), two demographic characteristics (age and gender), and domain knowledge of the storytelling performance being shown. Humor as the emotional excitation in vaudeville settings was rated using a 4-point (from 1 to 4) Likert scale. The humor rating scale (Nomura and Maruno, 2011) included four items that reflected the audience's degree of perceived humor (e.g., "I laughed or was inclined to laugh so much"). The transportive experience was rated using 18 items related to temporal transportation. Half of the items were derived from a subscale of the Literary Response Questionnaire (LRQ, Miall and Kuiken, 1995), which was translated into Japanese (Osanai and Okada, 2011). However, the wording of questions was inverted to fit into a vaudeville setting. For instance, "reading a novel" in the original text was changed to "viewing Rakugo" in the modified text (Nomura, 2013). The translated questionnaire also contained items relevant to the emotion of enjoyment in real life or items relevant to the author of the stories rather than transportation *per se* and less relevant items, which were not used. Moreover, as another aspect of the transportive experience, some items reflected subjective evaluation about participant's own somatic responses such as sweaty palms and chills (Nomura, 2013) were included. The questionnaire asked participants to write their age and gender in the blanks on the sheet and describe their knowledge of the story and the artist in the videotaped performance. The questionnaire also asked participants to describe their impressions of the performance. In addition, participants filled out information on their familiarity with Rakugo performances by (1) using media and (2) going to the theater in their everyday lives.

Procedure

The participants were separately invited to the laboratory room where the experimenter explained the experiment. To lessen the possibility that each participant intentionally controlled his/her eyeblink response, the experimenter withheld the actual purpose from the participants. Instead, the experimenter told the participants that the experiment "aimed to examine where you look on the monitor by measuring and recording the eye points while appreciating a Rakugo performance." After briefly explaining the eye-tracking device, a nine-point calibration was performed. The experimenter recorded the air temperature and humidity at the starting point and checked to ensure that the videos worked well. The experimenter played a video (a muted movie of fish swimming in a group), while measuring and recording the eyeblinks of each participant as an individual frequency and asynchrony baseline within each group. Each participant was then instructed how to play the movie and the experimenter left the room. Participants

started to play the assigned video on their own, while the device was measured and recorded their eyeblinks. After finishing the video presentation, the experimenter re-entered the room and asked the participants to complete the questionnaire. The experimenter explained that the actual purpose of the experiment was to measure the timings and frequency of eyeblinks while watching the storytelling performance. All participants gave permission for their eye information to be used in the study and agreed to answer the questionnaire. In addition, the experimenter asked if they had noticed that this was a study on eyeblinks. Five participants answered that they had noticed the eyeblink data, of which three were omitted from the analysis due to incomplete data (see, Participants). The other two participants suspected that the device might be related to eyeblink measurement; however, their eyeblink data were included in the analysis because they stated that they did not change the timing of their eyeblinks intentionally.
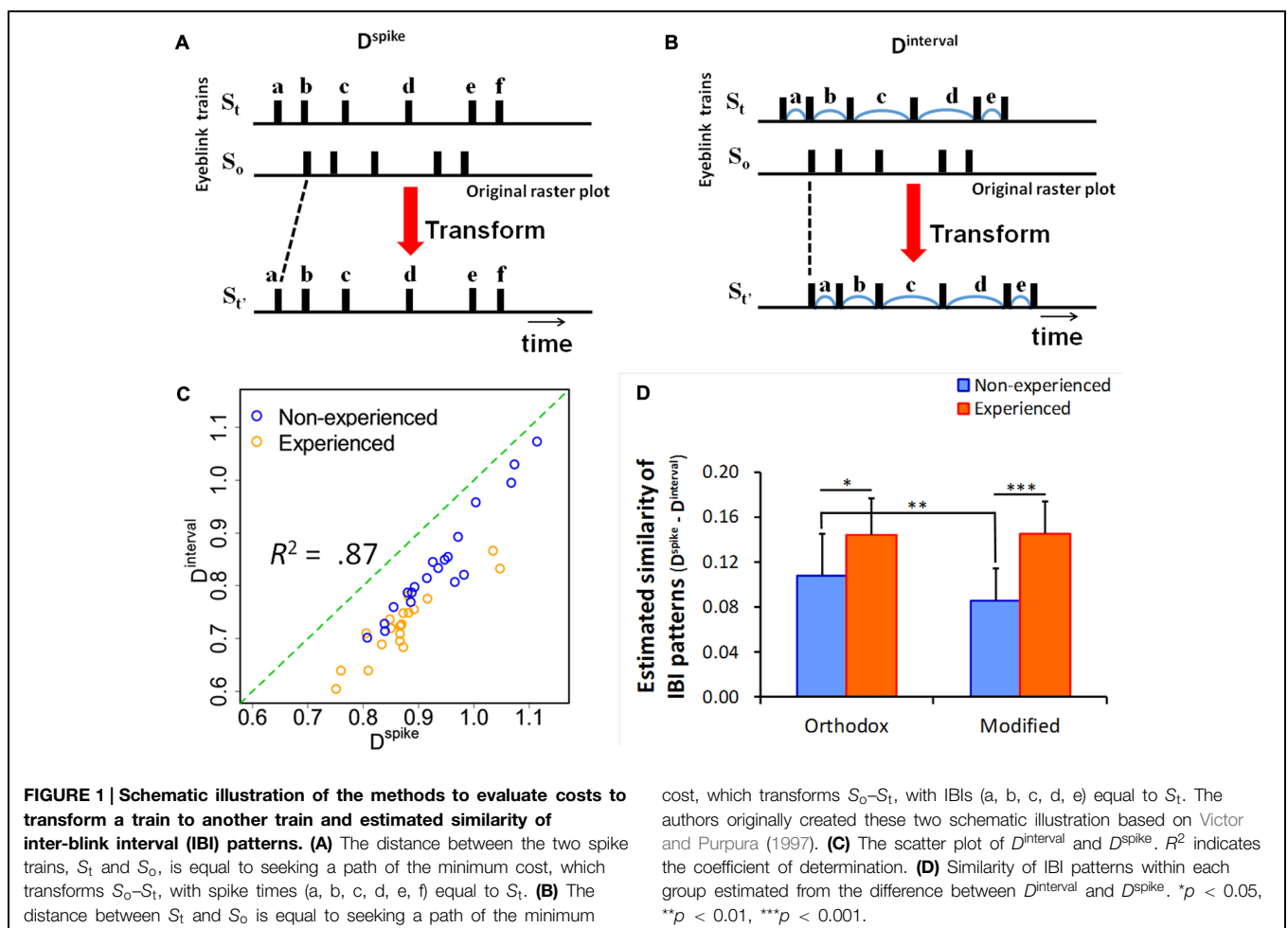
Analysis
Distance-Based Analysis of Blink (Spike) Trains: Asynchrony

Victor and Purpura (1997) proposed methods to quantify the asynchrony of two particular spike trains (e.g., time series of

intermittently firing neurons) focusing on the difference of spike timings. D^{spike} and $D^{interval}$ equally evaluate the distances of two different blinking trains (**Figure 1**). However, only D^{spike} calculates the distance at each time point of the spikes. In contrast, $D^{interval}$ takes into account the intervals of spike-by-spike. While these methods have been developed with the aim of analyzing asynchrony in firing neuron spike trains, they can be used to quantify the degree of asynchrony of particular blink trains.

D^{spike} is sensitive to inter-spike intervals. In contrast, $D^{interval}$ is sensitive to temporal spike patterns. Although a $D^{interval}$ value is constantly equal to or smaller than that of D^{spike}, there is no difference between the values of these indices if a particular temporal pattern is started at the same time. However, the value of $D^{interval}$ becomes smaller than that of D^{spike} when spike trains exhibit the same temporal pattern (motif) with a time delay in each time train (Victor and Purpura, 1997). Thus, the differences between these indices represent the degrees of pattern formation of IBIs. In other words, the difference in the value of $D^{interval}$ compared with that of D^{spike} suggests the ratio explained by the pattern similarity. If the viewing experience influences a situation model constructed through a viewing performance, a significant difference of pattern similarity will be found between experienced and non-experienced audiences.

FIGURE 1 | Schematic illustration of the methods to evaluate costs to transform a train to another train and estimated similarity of inter-blink interval (IBI) patterns. (A) The distance between the two spike trains, S_t and S_o, is equal to seeking a path of the minimum cost, which transforms S_o–S_t, with spike times (a, b, c, d, e, f) equal to S_t. **(B)** The distance between S_t and S_o is equal to seeking a path of the minimum

cost, which transforms S_o–S_t, with IBIs (a, b, c, d, e) equal to S_t. The authors originally created these two schematic illustration based on Victor and Purpura (1997). **(C)** The scatter plot of $D^{interval}$ and D^{spike}. R^2 indicates the coefficient of determination. **(D)** Similarity of IBI patterns within each group estimated from the difference between $D^{interval}$ and D^{spike}. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

In this study, the analysis unit was set to 250 ms because a blink usually occurs at an interval elapse of least 300 ms due to physiological limits (Nakano et al., 2009). That is, the whole video recording was divided into huge numbers of time windows (i.e., bins), each of which with a length of 250 ms, and the distance was counted based on the number of bins. To evaluate asynchrony of each scene during the performance, time trains of 5 min of performance time each (i.e., 1200 units = 4 bins/s × 60 s × 5 min) were used for calculation. As the total length was different with each video, the last 50 min of video footage was accurately divided into 10 scenes (i.e., each scene containing 5 min of footage). The rest (i.e., the first 22 s in the video for frequent viewers and 220 s in the video for first-time viewers) was excluded from the time-series analysis.

All values of D^{spike} and $D^{interval}$ were calculated using a modified version of a program provided by Victor and Purpura (1997). The program was mainly developed in the Matlab and Visual C++ environment. All p-values were two sided and a p-value of 0.05 or less was assumed statistically significant. All statistical analyses were performed using EZR (Easy R, Saitama Medical Center, Jichi Medical University; Kanda, 2013), which is a graphical interface for R (The R Foundation for Statistical Computing, Vienna, Austria, version 3.0.2).

Detecting the Onset of Laughter Elicited In *Situ*

To detect laughter, videos recorded in 30 frames per second were coded using ELAN 4.5.1 (Max Planck Institute for Psycholinguistics, Nijmegen), which has been developed for analyzing discourse processes and interactions among small-group members in face-to-face communication. The period of laughter was detected in the frame as the smallest unit (33.4 Hz) using only the sounds of the video. Each first frame was set as the onset of that laughter. A researcher trained in the methods of psychological study performed the coding procedure.

Results

Operational Checks
Laboratory Environment and Time Delay of Stimuli Presentation

No difference in the degree of laboratory humidity was found among the groups. Differences of time delays among the groups were not significant (range 501 ± 88.69 ms).

Audience Knowledge about the Performer and the Story

None of the non-experienced participants knew either the performer or the story. On the other hand, approximately a half of the experienced participants knew the performer (the orthodox performance: 46.67% and the modified performance: 33.33%) and the story (the orthodox performance: 66.67% and the modified performance: 46.67%; **Table 1**).

Reliabilities of Scales

The α coefficients of the scales were 0.77 and 0.91 for humor and experience of transportive experience, respectively. The coefficients were high enough for the following analysis.

TABLE 1 | Percentages of participants with knowledge of the performer and the story.

	Experienced participants	Non-experienced participants
Knowledge in advance of the performer		
Orthodox performance	46.67	0.00
Modified performance	33.33	0.00
Knowledge in advance of the story		
Orthodox performance	66.67	0.00
Modified performance	46.67	0.00

n = 15 for each group.

Baseline of Asynchrony

To confirm that there was no difference in the total count of eyeblinks per time between groups, ANOVA was used for the IBI expressions of performer and viewing experience of the audience. The results showed no main effect and no interaction. Thus, the total rates or total numbers of eyeblinks were not different among the groups. This result was supported even if the participant's age—a factor that may have influenced the total numbers of eyeblinks—was taken into account. Under the baseline condition, only the main effect of audience viewing experience was significant [$F(1,338) = 9.84$, $p < 0.01$]. The experienced audience value of $D^{interval}$ was lower than that of the non-experienced audience (0.62 vs. 0.70, $p < 0.05$ for orthodox video and 0.60 vs. 0.71, $p < 0.10$ for modified video, respectively). This fact may indicate that experienced audience members slightly tend to synchronize their eyeblinks even when they are watching a video unrelated to domain knowledge (silent movie of a group of fish). Owing to this result, in Section "Temporal Pattern of $D^{interval}$," differences between the value of $D^{interval}$ for experienced and that for non-experienced audiences were accepted only when the effect size of this comparison exceeded that of the baseline, and statistical values were significant. The values of asynchrony under the baseline condition in each group were relatively lower than those during video screening. Because the stimulus used in the baseline condition contains only visual information, the timing of the allocation would converge. On the other hand, storytelling performances included audio-visual stimulus requiring participants to integrate multimodal information.

Multiple Regression Analysis

The mean and SD of IBIs followed logarithmic normal distribution according to the nature-of-time relevant variable. In the following analysis, logarithmic-transformed mean and SDs of IBIs were used. To examine the relationships between variables, zero-order correlations were calculated (**Table 2**). The coefficient of correlation between transportive experience and humor was very high ($r = 0.772$, $p < 0.001$). Knowledge of the performer positively correlated with transportive experience and humor. The *SD* of IBIs did not have a salient correlation with other variables.

In order to explore which variables predicted the experience of transportation, a multiple regression analysis was performed (**Table 3**). The results of multiple regression analysis demonstrated that humor strongly predicted the experience of

TABLE 2 | Zero-order correlation coefficients between variables used for multiple regression analysis.

	2	3	4	5	6
1. Transportation	0.772***	0.071	0.197†	0.262*	0.107
2. Humor	–	−0.008	−0.014	0.315**	0.191†
3. Mean of inter-blink intervals (IBIs)		–	0.156	0.033	−0.028
4. SD of IBIs			–	0.085	0.105
5. Knowledge of the performer				–	0.703***
6. Knowledge of the story					–

*N = 60, †p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.*

TABLE 3 | Regression analysis of humor ratings, IBI statistical values, and domain knowledge of experience of transportation.

	B	SE	Standardized β	*t*-test
Intercept	0.362	0.726		0.499
Humor	0.503	0.056	0.772	9.064***
Mean of IBIs	0.047	0.101	0.038	0.462
SD of IBIs	0.432	0.171	0.208	2.525*
Knowledge ofthe performer	0.102	0.140	0.085	0.724
Knowledge ofthe story	−0.128	0.121	−0.121	−1.059

*R = 0.806, R^2 = 0.649, *p < 0.05, ***p < 0.001.*

transportation ($\beta = 0.772$, $p < 0.001$). *SD* of IBIs also regressed on the experience of transportation ($\beta = 0.208$, $p < 0.05$). The other variables such as age, gender, means of IBIs, and domain knowledge (the performer and the story, dummy variables) exhibited no significant effects. The zero-order correlations between the domain knowledge and transportive experience were weakened by taking the other variables into consideration. The coefficient of determination was considerably high ($R^2 = 0.64$).

Experience of Transportation

To reveal the effect of the performer's expressions and audience viewing experience, the factor design was a two-way ANOVA performance (between the 2nd level; for frequent viewers and first-time viewers) × experience of audience (between the 2nd level; experienced and non-experienced). The dependent variables analyzed included the score of humor scale and the score of the transportation scale. When humor and transportation scale scores were combined, the main effects and interaction between performance and experience of audience were not significant. For the transportation scale score, the effect of experience was marginally significant, indicating that the score of the experienced audiences was very slightly higher than that of the non-experienced [$F_{(1,56)} = 2.963$, $p < 0.10$, experienced 2.58 vs. non-experienced 2.37]. In addition, the simple main effect (corrected Bonferroni's method) of the experienced audiences was marginally significant (0.30, SE $= 0.17$, $p < 0.10$, experienced 2.69 vs. non-experienced 2.39).

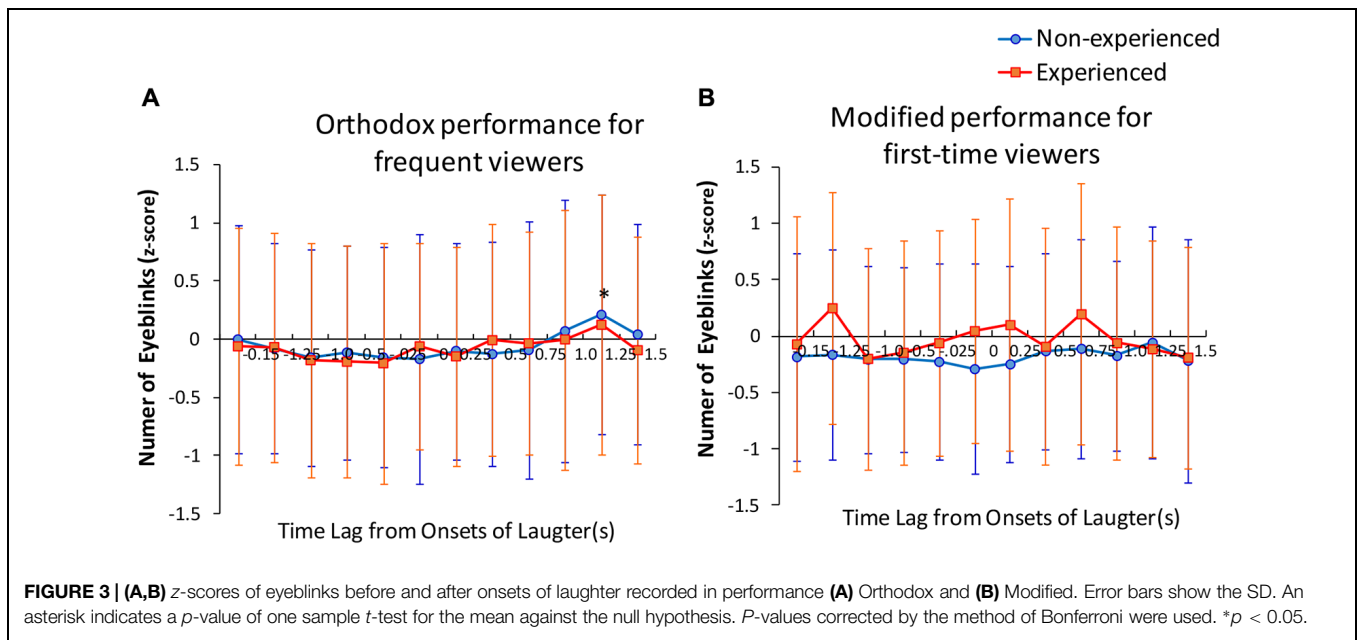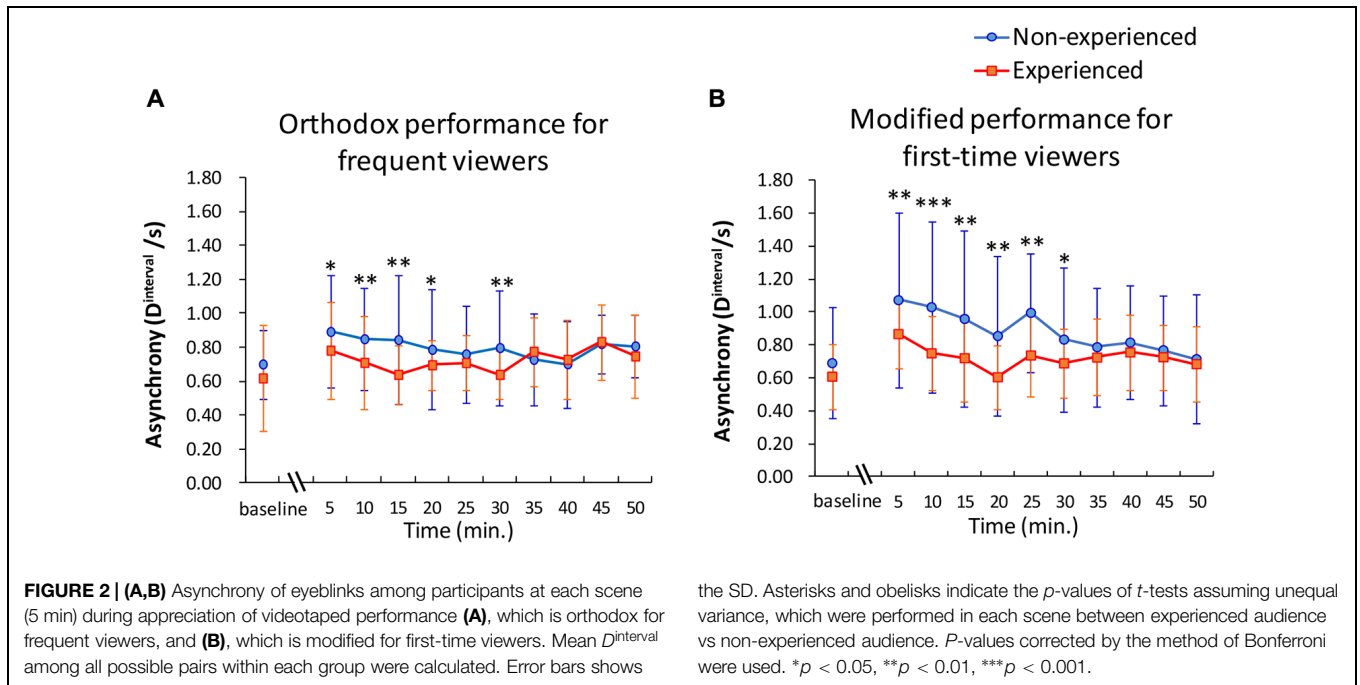## Eyeblink Synchronization
Estimated Similarity of IBI Patterns

The difference between D^{interval} and D^{spike} indicates the similarity of the IBI patterns within two trains (**Figures 1A,B**). As the results of the two-way ANOVA (viewing experience × video) data of 10 scenes showed, the main effect of the viewing experience was significant [$F_{(1,2066)} = 25.38$, $p < 0.0001$, **Figure 1D**]. Sub-effect tests revealed that the estimated similarity in eyeblinks of the experienced audience was higher than that of the non-experienced audience in both performances ($p < 0.001$, 0.15 vs. 0.09 and $p < 0.05$, 0.14 vs. 0.11). In addition, the estimated similarity in eyeblinks during the orthodox video was higher than that of the modified video ($p < 0.01$, 0.12 vs. 0.10).

Temporal Pattern of D^{interval}

In the first six scenes (0–30 min) of each video, the average D^{interval} within a group of experienced participants was significantly higher than that of the non-experienced participants, with the exception of 25 min of the orthodox performance. All of the effect sizes of these comparisons exceeded those of the comparisons observed under the baseline condition. Regarding homogeneity of variances, the null hypothesis that the true ratio of variances is equal to 1 was rejected at D^{interval} from 15 to 30 min of the orthodox video and during all D^{interval} of the modified video (not shown in **Figure 2**). Overall, the index D^{interval} of the participants who had viewing experience remained low while watching both the orthodox video (**Figure 2A**, orange line) and the modified video (**Figure 2B**, orange line). Non-experienced participants who watched the modified performance gradually reduced eyeblink asynchrony as the story developed (**Figure 2B**, blue line). On the other hand, even non-experienced participants had reduced asynchrony as of the first few minutes to the end of the story while watching the orthodox video (**Figure 2A**, blue line). The SDs for experienced participants also stayed relatively small while SDs for non-experienced participants decreased throughout the performance.

Effect of Laughter

In the case of the orthodox performance, the results of above mentioned ANOVA demonstrated relatively lower levels of D^{interval} for even non-experienced participants as of the beginning of the performance. However, these effects for non-experienced participants were not observed in the modified performance, possibly because of the differences between audience responses *in situ* reflecting changes in the emotional expression of the performance. In order to reveal the possible influence of laughter on the difference in D^{interval}, the number of eyeblinks occurring at, before, and after the onset of laughter was compared. A three-way ANOVA was performed on the number of eyeblinks in each unit was normalized to a z-score across the performances (**Figure 3**). The factor design of this ANOVA was video (between the 2nd level; orthodox and modified) × experience of audience (between the 2nd level; experienced and non-experienced) × timing (between the 12th level; the time of six units before onset and six units after onsets of laughter). First, Mauchly's test was conducted to check sphericity. All of the test statistics were not significant. We then used type III

FIGURE 2 | (A,B) Asynchrony of eyeblinks among participants at each scene (5 min) during appreciation of videotaped performance **(A)**, which is orthodox for frequent viewers, and **(B)**, which is modified for first-time viewers. Mean $D^{interval}$ among all possible pairs within each group were calculated. Error bars shows the SD. Asterisks and obelisks indicate the p-values of t-tests assuming unequal variance, which were performed in each scene between experienced audience vs non-experienced audience. P-values corrected by the method of Bonferroni were used. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

FIGURE 3 | (A,B) z-scores of eyeblinks before and after onsets of laughter recorded in performance **(A)** Orthodox and **(B)** Modified. Error bars show the SD. An asterisk indicates a p-value of one sample t-test for the mean against the null hypothesis. P-values corrected by the method of Bonferroni were used. *$p < 0.05$.

sum of square repeated measures ANOVA assuming sphericity. The main effects of time were significant [$F(11,3916) = 1.853$, $p < 0.05$) and none of the other main effects and interactions were significant. To identify the sub-effect, a two-way ANOVA for each video (orthodox and modified) was exerted. The results demonstrated that the main effect of timing was significant during the orthodox video [$F(11,2420) = 2.21$, $p < 0.05$). We performed one sample t-test for the mean against the null hypothesis (μ is 0) using a p-value collected by the Bonferroni's method. Only a time point 1.25–1.50 s after the onset of laughter in the video was significantly higher than 0. All means at the other time points were not significant for rejecting the null hypothesis.

Discussion

Mechanisms of Transportation

Participants' eyeblinks synchronized among the non-experienced audience at a level equivalent to that of the experienced participants through an appreciation of the performance (**Figure 2A**). As enough information seemed to be presented in each

performance, even non-experienced participants appeared to be able to construct a situation model using only a temporally accumulated knowledge of the story by comprehending the storyline and the personalities of the characters. On the other hand, the SDs of the experienced participants tended to be lower than those of the non-experienced participants. This result suggests that the audience's domain knowledge cultivated by viewing experience aids in the construct of similar situation models among the audience. The results of estimated similarity of the IBI pattern (**Figure 1C**) also suggest that experienced audiences, compared to non-experienced audiences, respond in more reproducible ways within each group. Although not all experienced participants knew the story or the performer perfectly, the experience of the participants helped to synchronize their eyeblinks. Thus, results were obtained by application of knowledge regarding typical developing patterns of storylines in the field of Rakugo performance.

However, in this experiment, the situation model supported by domain knowledge did not explain an experience of transportation fully (**Figure 1D**). The results of the ANOVA concerning transportation showed that the main effect of audience experience was weak. The results of the multiple regression analysis indicated that humor and SD of IBIs predicted a transportive experience. Other variables had no predictive effects. Van Laer et al. (2014) reveals that age, gender, and knowledge gained by education, among other variables, affects the degree of a transportive experience, based on a review of several articles (e.g., Stern, 1992; Green and Brock, 2000; Diekman and Murnen, 2004). However, the apparent effects relating to the degree of viewing experience and other demographic variables seem to be peripheral. The SD of IBIs suggests that an individual's allocation of attention varies more frequently as he or she is inclined to predict upcoming events (Nomura and Okada, 2014). It could be said that the eyeblink-rate variability is accompanied by emotional excitement. This emotionally motivated eyeblink-rate variability might be attributed to the expressiveness of a performance and corresponding humor *in situ*. Because the same story was performed by the same performer, the differences of asynchrony must depend on the performance rather than the structure of the story. As described so far, the two mechanisms that we mentioned earlier seemed to be confirmed. It was suggested that the emotionally excited eyeblink-rate variability could be a good predictor of transportation.

The possibility of eyeblink occurrence increased at 1.25–1.50 s after the onset of laughter. This result suggests that laughter by the surrounding audiences functions as a cue for further processing (Nomura and Maruno, 2008). A time delay from the onset of laughter may be due to a time lapse between recognition and reinterpretation of a situation in the story. However, the effect was confirmed only during the orthodox performance. For non-experienced audiences, estimated pattern similarity (i.e., formation ratio of temporal patterns, "motif") was also higher for those who watched the orthodox video than those who watched the modified video, as shown in **Figure 1D**. These results suggest that even non-experienced audiences synchronize their eyeblinks, to some extent, when appreciating a performance acted in the orthodox way usually seen in theaters.

The performance that amuses experienced audiences would seem to simultaneously exert this effect on non-experienced audiences.

Although the effect of the viewing experience was confirmed, it was weak. A non-experienced audience might devote significant cognitive resources to comprehending the contents of the story, leaving very little for other resources. In contrast, an experienced audience might be engaged in a transportive experience by sparing cognitive resources in order to appreciate the details of expression, especially for an orthodox performance. The experienced audience might sometimes pay attention to a particular nuance of expression by each artist rather than simply enjoy the contents of the performance *per se*. Actually, in the free description about their impressions of the performance, some experienced audience members answered that the performer appeared to inherit the traditional style of Rakugo compared to the other performers in his generation. A viewing experience does not always lead to transportation. An implicit selection of information and a resupply of emotionally excited attention lead to a precise prediction of the next plot twist and an engrossing experience. Overall, a transportive experience would actualize under a situation in which both active leading by performance and active anticipating by the audience occur. In this sense, a performer and audience share the responsibility to create transportive enjoyment in a vaudeville setting. A performer would act as the leader in providing his/her creative expressions and the audience would play the role of actively anticipating the created world of the story.

Dynamic Indices and Future Direction

The findings about the underlying mechanisms in real-time processing are significant in the research field of transportive experience that has focused on the traits of the receiver (see Van Laer et al., 2014). In particular, the predictive power of eyeblink-rate variability during viewing performance implies that people experience transportation through active coordination of specific external audio-visual information. Both eyeblink synchronization and eyeblink-rate variability could be useful measurements for researchers to infer the inner experience of audience members by observing unintentional behaviors objectively. The results of this study also suggest that emotional excitement motivates more attentional cognitive resources onto the actor's expressions and the structure of the story. In this study, the positive emotion (i.e., humor) is strongly related to the transportive experience because the performance is oriented to create a sense of enjoyment or exhilaration in the audience in a vaudeville setting. However, it is not surprising if the feelings of thrill or suspense predict a transportive experience at the cinema. Future research is necessary to examine the relationships between excitement of other kinds of emotion and transportive experiences.

A possibility exists that transportation is weakened compared with that experienced through live performance because a videotaped performance cannot preserve the atmosphere *in situ*. Further study is necessary to clarify whether or not emotionally excited eyeblink-rate variability more strongly facilitates the transportive experience in real vaudeville settings. Although the

humor experience was evaluated retrospectively owing to operational limitations in this study, future research will reveal the time-sequential relationships between transportation, emotional excitement, and eyeblink-rate variability by measuring ongoing physiological indices such as skin conductance, heart rate, and aspiration rhythm.

Acknowledgments

References

Diekman, A., and Murnen, S. (2004). Learning to be little women and little men: the inequitable gender equality of nonsexist children's literature. *Sex Roles* 50, 373–385. doi: 10.1023/B:SERS.0000018892.26527.ea

Green, M. C., and Brock, T. C. (2000), The role of transportation in the persuasiveness of public narratives. *J. Pers. Soc. Psychol.* 79, 701–721. doi: 10.1037/0022-3514.79.5.701

Hall, A. (1945). The origin and purposes of blinking. *Br. J. Ophthalmol.* 29, 445–467. doi: 10.1136/bjo.29.9.445

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness* (No. 6). Cambridge, MA: Harvard University Press.

Kanda, Y. (2013). Investigation of the freely-available easy-to-use software "EZR" (Easy R) for medical statistics. *Bone Marrow Transplant* 48, 452–458. doi: 10.1038/bmt.2012.244

Mainen, Z. F., and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science* 268, 1503–1506. doi: 10.1126/science.7770778

Miall, D. S., and Kuiken, D. (1995). Aspects of literary response: a new questionnaire. *Res. Teaching English* 29, 37–58.

Nakano, T., and Kitazawa, S. (2010). Eyeblink entrainment at breakpoints of speech. *Exp. Brain Res.* 205, 577–581. doi: 10.1007/s00221-010-2387-z

Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., and Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. *Proc. R. Soc. B Biol. Sci.* 276, 3635–3644. doi: 10.1098/rspb.2009.0828

Nomura, R. (2013). Construction of Scale of transportation experience during appreciating rakugo performance. *RIDEO Jpn. J. Laughter Humor Stud.* 20, 32–43.

Nomura, R., and Maruno, S. (2008). An integration of humor generation theories and a proposal of dynamical comprehension and elaboration theory. *Jpn. Psychol. Rev.* 51, 500–525.

Nomura, R., and Maruno, S. (2011). Constructing a coactivation model for explaining humor elicitation. *Psychology* 2, 477–485. doi: 10.4236/psych.2011.25074

Nomura, R., and Okada, T. (2014). Spontaneous synchronization of eyeblinks during story-telling performance. *Cogn. Stud.* 21, 226–244. doi: 10.11225/jcss.21.226

Osanai, K., and Okada, H. (2011). Construction of the literary response questionnaire of Japanese. *Shinrigaku Kenkyu* 82, 167–174. doi: 10.4992/jjpsy.82.167

Radvansky, G. A., and Copeland, D. E. (2001). Working memory and situation model updating. *Memory Cogn.* 29, 1073–1080. doi: 10.3758/BF03206375

Sestir, M., and Green, M. C. (2010). You are who you watch: identification and transportation effects on temporary self-concept. *Soc. Influ.* 5, 272–288. doi: 10.1080/15534510.2010.490672

Stern, B. B. (1992). Feminist literary theory and advertising research: a new "reading" of the text and the consumer. *J. Curr. Issues Res. Advert.* 14, 9–21. doi: 10.1080/10641734.1992.10504976

Van Dijk, T. A., and Kintsch, W. (1983). *Strategies of Discourse Comprehension.* New York, NY: Academic Press.

Van Laer, T., de Ruyter, K., Visconti, L. M., and Wetzels, M. (2014). The extended transportation-imagery model: a meta-analysis of the antecedents and consequences of consumers' narrative transportation. *J. Consum. Res.* 40, 797–817. doi: 10.1086/673383

Victor, J. D., and Purpura, K. P. (1997). Metric-space analysis of spike trains: theory, algorithms and application. *Network* 8, 127–164. doi: 10.1088/0954-898x/8/2/003

Zwaan, R. A., Magliano, J. P., and Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *J. Exp. Psychol. Learn Mem. Cogn.* 21, 386–397. doi: 10.1037/0278-7393.21.2.386

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Losing track of time through delayed body representations

Thomas H. Fritz[1,2,3]*, Agnes Steixner[1], Joachim Boettger[1] and Arno Villringer[1]

[1] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [2] Department of Nuclear Medicine, University of Leipzig, Leipzig, Germany, [3] Institute for Psychoacoustics and Electronic Music, Ghent, Belgium

The ability to keep track of time is perceived as crucial in most human societies. However, to lose track of time may also serve an important social role, associated with recreational purpose. To this end a number of social technologies are employed, some of which may relate to a manipulation of time perception through a modulation of body representation. Here, we investigated an influence of real-time or delayed videos of own-body representations on time perception in an experimental setup with virtual mirrors. Seventy participants were asked to either stay in the installation until they thought that a defined time (90 s) had passed, or they were encouraged to stay in the installation as long as they wanted and after exiting were asked to estimate the duration of their stay. Results show that a modulation of body representation by time-delayed representations of the mirror-video displays influenced time perception. Furthermore, these time-delayed conditions were associated with a greater sense of arousal and intoxication. We suggest that feeding in references to the immediate past into working memory could be the underlying mental mechanism mediating the observed modulation of time perception. We argue that such an influence on time perception would probably not only be achieved visually, but might also work with acoustic references to the immediate past (e.g., with music).

Keywords: time perception, body schema, visual delay, virtual mirror, flow, intoxication

Introduction

Body representation and sense of time in humans seem to be closely related. Previous research showed that several physiological parameters such as heart rate (Wittmann, 2013) and body temperature (Wearden and Penton-Voak, 1995) modulate the way we perceive time. Also, the mental representation of one's own body, the body schema, seems to play an influential role in the subjective perception of time. Body schema has been defined as the "mental construct that comprises the sense impressions, perceptions, and ideas about the dynamic organization of one's own body and its relations to that of other bodies" (Berlucchi and Aglioti, 1997). The perception of body schema can be modified by a variety of means (Holmes and Spence, 2006). An established method to achieve this is in an experimental setting where visual feedback is presented via mirrors or video monitors (Holmes and Spence, 2004; Bernardi et al., 2013).

Such manipulations of body schema can be employed to create proprioceptive/sensory changes and illusions such as, for example, the rubber hand illusion (Holmes and Spence, 2007), and out-of-body experiences (Lenggenhager et al., 2007). A manipulation of body schema, however, also seems to relate to recreational experiences that are associated with an altered perception of time. For example, the consumption of cannabis has been shown to affect both bodily representation

and sense of time (Edelmann and Terbuyken, 2002; Atakan et al., 2012). Methods not related to intoxication, such as meditating, that employ changes in the monitoring of body schema have also been shown to have—among other psychological effects—an effect on the perception of time (Berkovich-Ohana et al., 2013). Likewise, the subjective experience of flow (a psychological state associated with high motivation and either motor or cognitive challenge) that often corresponds to a change in bodily experience appears to be associated with a modified perception of time (cf. Jackson and Csikszentmihaly, 1999).

Further evidence for a relation of body schema and time perception derives from schizophrenia research. Schizophrenia patients, who have been found to display a different body representation schema (Thakkar et al., 2011), showed greater variability with regard to their performance in a timing task and impaired ability to estimate time (Peterburs et al., 2013; Bolbecker et al., 2014).

Taken together, previous findings strongly suggest an association of one's own body representation and one's perception of time, so that a modulation of own body representation might affect the ability to accurately estimate time. In order to further investigate this relationship, we designed an experiment in which we manipulated bodily representations via delayed and non-delayed virtual mirrors. Because variables such as sex and age have been shown in the past to be associated with the perception of time, they were assessed to identify possible influences as well as moderator or mediator effects (Perbal et al., 2002; Glicksohn and Hadad, 2012).

Materials and Methods

Participants

Seventy individuals (39 females, age 30.01 ± 14.21, range: 11–67) took part in the experiment. Required minimal age to take part in the study was 10 years. Participants gave their written consent and did not receive any compensation.

Procedure

Each participant was randomly assigned to one of four groups. There were two experimental groups (aware delayed and non-aware delayed) and two control groups (aware not delayed and non-aware not delayed). We used a 2x2 design with delayed videos vs. non-delayed videos of body representations as one factor, and time-awareness vs. no time-awareness[1] as the other. In both the delay conditions, participants were presented with visual delayed videos of themselves, while in the two non-delayed groups videos were shown in real-time. Individuals were additionally assigned to one of the two "time-awareness" conditions (aware delay: $n = 19$; aware no delay: $n = 20$). In the awareness conditions they were instructed to enter the installation and return after they assumed 90 s to be over. Participants in the "no

time-awareness" conditions (not aware delay: $n = 16$; not aware no delay: $n = 15$) were asked to enter the installation and stay there as long as they wanted. Prior to the experiment participants were instructed to fill out a questionnaire that also included constructs such as Situational Self-Awareness, Attentional Resource Allocation, Time Perspective, and Impulsiveness. Because the current paper aims to focus on a relation of body representation and time perception, results of these measurements are not reported here. Afterward, participants were asked to leave their timekeepers (i.e., cell phones, watches) with the experimenter for the duration of their stay in the experimental installation (**Figure 1**), allegedly in order to avoid distraction.

How long each participant stayed in the installation was unobtrusively measured by the experimenter. All participants received a post-experimental questionnaire including measures of arousal, perceived intoxication, subjective experience of changes in time, and several other scales not relevant to the current study, as well as some demographic questions.

A video signal was recorded by a camera installed above one of the eight video screens. This signal was then transferred to a media-server computer. A software (Wings Vioso, Duesseldorf, Germany) allowed to systematically delay the video signal and output eight different delayed versions (1–4 s) of the video signal during the delay condition. Groups of two outputs were transmitted to one of four beamers that projected the video signals on back projection screens (one on each of two screens), so that the participants could see the eight different visual presentations on all eight screens without artifacts (e.g., shadows).

Data Analysis

Estimation error was computed in the awareness conditions as the measured duration of stay in the installation minus 90 s and in the no awareness conditions as the measured duration of stay minus the estimate of duration of stay by participants. Similar to previous research on human time estimation or production, we used absolute estimation error as an indicator of magnitude of estimation error, which is irrespective of its direction (Pastor et al., 1992; Rammsayer, 1997; Barkley et al., 2001a,b; Kerns et al., 2001). Further, following the procedures of Bauermeister et al. (2005), we applied logarithmic transformation on absolute estimation errors, i.e., estimation error plus 1 was computed and log-transformed, to reduce skewness of our data and to bring the raw absolute values more in line with the assumptions of parametric statistics. Because it was not possible to directly investigate differences between signed estimation errors due to properties of data distribution, we computed transformed signed estimation errors by adding the highest negative estimation error (195 s) to get positive values. In order to reach a normal distribution these values were squared afterwards. In order to compute effects on the direction of estimation errors, estimation errors greater than 0 s were coded as an overestimation of duration, and an error below 0 s was coded as an underestimation of duration.

The variables age and duration of stay were split into two sections by their median, i.e., younger vs. older age and long vs. short duration of stay. The cut-off for high vs. low arousal was defined as the value of three on the administered 6-point scale.

---

[1]Note that "time awareness" should be understood as relative to the non-awareness conditions, i.e., pointing to increased attention dedicated to time by participants compared to participants assigned to the no time-awareness conditions. This is due to the fact that we did not employ a control task that made sure that participants were constantly focusing on the passage of time in the time-awareness conditions.
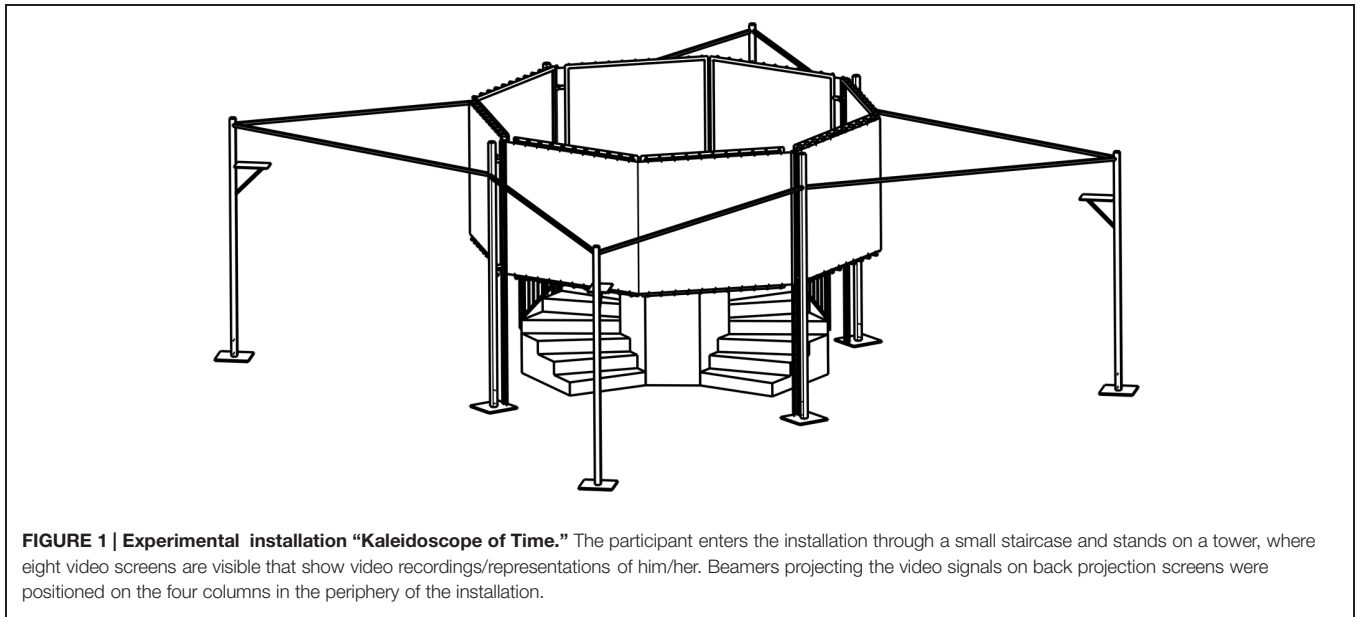
FIGURE 1 | Experimental installation "Kaleidoscope of Time." The participant enters the installation through a small staircase and stands on a tower, where eight video screens are visible that show video recordings/representations of him/her. Beamers projecting the video signals on back projection screens were positioned on the four columns in the periphery of the installation.

Parametric computations (i.e., ANOVA, independent samples *t*-test) were used when requirements were met by the data. If basic assumptions were violated, non-parametric computations (i.e., Kruskal–Wallis test, Wilcoxon Signed-Rank test, *U*-test) were employed.

Measures

Arousal

Experienced arousal was measured by means of the one-item Felt Arousal Scale (FAS; Svebak and Murgatroyd, 1985). Participants had to indicate their arousal on a 6-point scale with 1 indicating "low arousal" and six indicating "high arousal." This measure of arousal or activation was found to correlate with other arousal scales ranging from 0.45 to 0.70, and thus showed convergent validity (Sheppard and Parfitt, 2008).

Intoxication

Participants indicated their perception of intoxication assessed by the question "Wie 'berauscht' fühlen Sie sich im Moment?" ("How 'intoxicated' do you feel at the moment?") on a 4-point Likert scale with 1 indicating "not at all" and four indicating "very much."

Subjective Changes in Sense of Time

Subjectively experienced changes in sense of time were assessed by the subdimension of "time sense" of the Phenomenology of Consciousness Inventory (PCI; Pekala et al., 1986). The overall measure was shown to have sufficient reliability and validity for assessing changes in phenomenological experience (Pekala et al., 1986). The subdimension used in this study consists of three items, each comprising two opposing statements. Participants rated the statements on a 5-point scale, indicating with which of the two statements they agreed more, e.g., a rating of three indicating equal agreement with both statements. Sample item: "Time seemed to greatly speed up

or slow down" vs. "Time was experienced with no changes in its rate of passage." In our study, Cronbach's alpha was $\alpha = 0.83$.

Other Outcomes and Controls

The experiences of feeling detached from one's body and of "sinking into pictures" were assessed by single items constructed for the purpose of this study. These items were rated on a 4-point Likert scale ranging from "not at all" to "completely." For the assessment of subjectively experienced waste of time the item "In your honest opinion, do you think this survey was a waste of time?" was used (O'Brien et al., 2011). In addition, participants in the time-awareness conditions were asked if they used counting to estimate time correctly. Finally, data on age and sex were collected.

Results

We constructed a general linear model (GLM), including the variables *group*, *duration of stay* (long vs. short), *age* (younger vs. older) and *sex*. In addition, we modeled interactions between *sex × group*, *age × group*, and *duration of stay × group*. *Group* was found to have a significant effect on log-transformed estimation error, $F(3,53) = 3.36$, $p = 0.025$, $\eta^2 = 0.160$. Moreover, a significant interaction existed between duration of stay and group, $F(3,53) = 4.67$, $p = 0.006$, $\eta^2 = 0.209$. No significant main effects were observed for sex, $F(1,53) = 3.13$, $p = 0.083$, $\eta^2 = 0.056$, duration of stay, $F(1,53) = 0.08$, $p = 0.781$, and age, $F(1,53) = 0.05$, $p = 0.831$. Further no significant interactions existed between sex and group, $F(3,53) = 1.20$, $p = 0.319$ or age and group, $F(3,53) = 0.83$, $p = 0.485$. The overall corrected model accounted for 42.8% of variance, $F(15,53) = 2.65$, $p = 0.005$ (for mean and standard deviation see **Table 1**).

Due to the non-normality of data, $W(69) = 0.945$, $p = 0.005$, and inhomogeneity of variances, no ANOVA or Kruskal–Wallis

TABLE 1 | Mean and standard deviation of absolute time estimation errors (log-transformed and non-transformed) for experimental and control groups.

Experimental group	Log-transformed absolute estimation error			Absolute estimation error in s		
	Duration of stay					
	Short	Long	Total	Short	Long	Total
Aware delay	1.41 (0.23)	0.95 (0.51)	1.24 (0.42)	28.50 (16.03)	13.29 (14.40)	22.89 (16.82)
Aware no delay	1.49 (0.24)	1.34 (0.46)	1.42 (0.35)	33.64 (15.97)	34.33 (35.91)	33.95 (26.03)
Not aware delay	1.31 (0.37)	1.85 (0.39)	1.61 (0.46)	24.71 (15.86)	90.00 (56.30)	61.44 (53.95)
Not aware no delay	1.29 (0.28)	1.50 (0.31)	1.38 (0.30)	23.00 (18.59)	37.83 (29.93)	29.36 (24.26)
Overall	1.39 (0.27)	1.43 (0.52)	1.41 (0.40)	28.13 (16.38)	46.59 (47.59)	36.35 (35.06)

test could be performed to detect possible effects of group on means of estimation error. A Welch test computed on the basis of the transformed signed estimation error did not yield any significant differences between groups, $W_{(3,32)} = 0.690$, $p = 0.565$. However, the descriptive pattern of group means of estimation error is shown in **Figure 2**. The overall mean of estimation error was $M = -12.75$, $SD = 49.04$. A Kruskal–Wallis test did not yield any differences on the duration of stay between groups, $\chi^2 = 1.880$, $p = 0.598$.

No significant effect was found of groups on the direction of estimation error, i.e., overestimation vs. underestimation, $\chi^2_{(3}$, $n = 69) = 6.433$, $p = 0.092$[2]. Furthermore, a Chi-Square test revealed neither significant differences between gender and direction of estimation error, $\chi^2_{(1}$, $n = 69) = 0.122$, $p = 0.727$, nor between age groups and direction of estimation error (age < 25 vs. age > 24), $\chi^2_{(1}$, $n = 69) = 0.349$, $p = 0.555$. Additionally, an ANOVA yielded no effect of perceived waste of time (three

groups) by subjects on log-transformed absolute estimation error, $F_{(2,66)} = 0.454$, $p = 0.637$.

With regard to subjective changes in the perception of time, we found that groups differed significantly on the degree of perceived changes in time sense, $\chi^2_{(3}$, $n = 70) = 8.674$, $p = 0.034$. According to ranks, participants of the *aware delayed group* experienced the strongest subjective changes in the passage of time, followed by the *aware no delay* and *not aware delay* groups. Participants in the *not aware no delay* group reported the lowest changes in time sense.
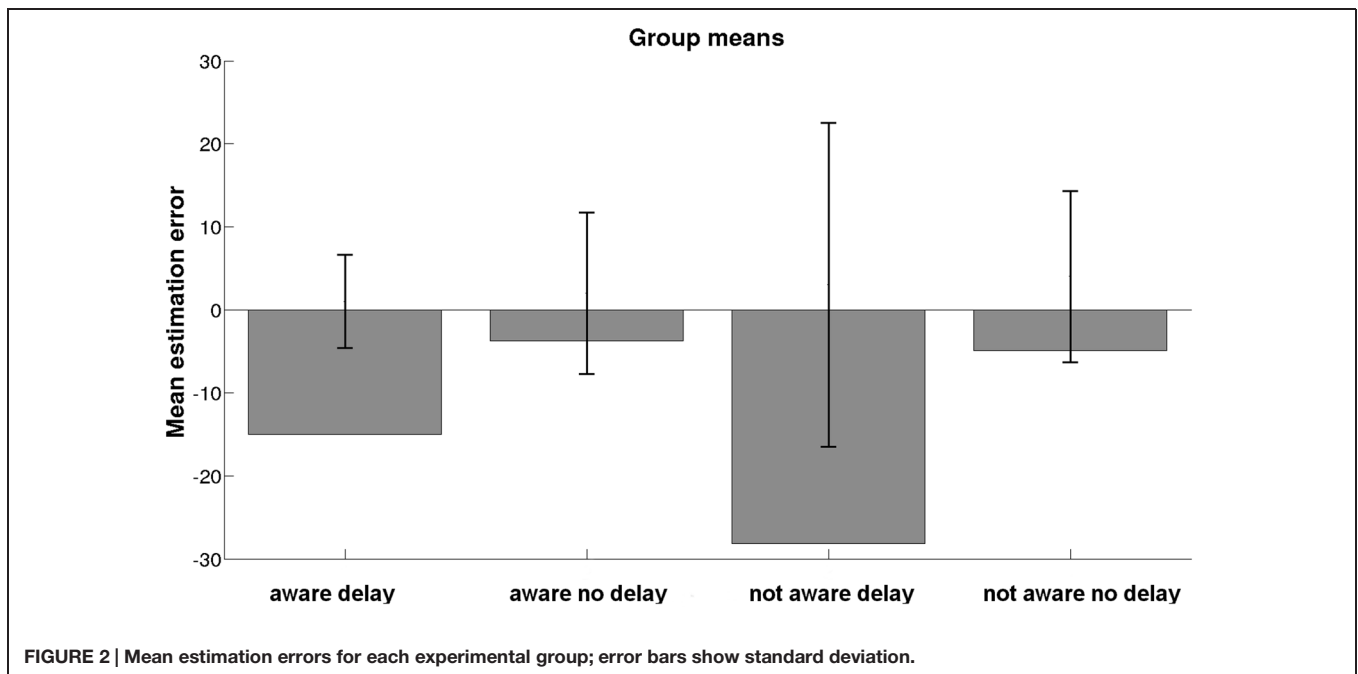
Further, a Mann–Whitney U-test showed a significant effect of delay condition (delay or no delay) on the feeling of arousal, $U = 387,500$, $p = 0.009$[3], on perceived intoxication, $U = 338.500$, $p = 0.013$[4], and a marginal effect on the experience of 'sinking into the pictures,' $U = 444.500$, $p = 0.038$[5]. It did not have an effect on feeling detached from one's body, $U = 468.000$,

[3]Bonferroni–Holm corrected alpha level: α' = 0.0125.

[4]Bonferroni–Holm corrected alpha level: α' = 0.0167.

[5]Bonferroni–Holm corrected alpha level: α' = 0.025.

[2]Bonferroni–Holm corrected alpha level: α' = 0.0167.

FIGURE 2 | Mean estimation errors for each experimental group; error bars show standard deviation.

$p = 0.071$. With regard to the experimental 'awareness' condition (when participants were instructed to stay in the installation for 90 s), no significant effects on arousal, $U = 492.000$, $p = 0.219$, perceived intoxication, $U = 483.000$, $p = 0.873$, 'sinking into pictures,' $U = 596.000$, $p = 0.916$, feeling detached from one's body, $U = 495.500$, $p = 0.171$[6], were found. A Kruskal–Wallis test did not yield any differences on the duration of stay between groups, $\chi^2(3, n = 70) = 1.880$, $p = 0.598$.

Discussion

Here, we demonstrate an influence of temporally manipulated visual own-body representations on time perception and a number of other psychological parameters. A delayed video presentation of one's own body image resulted in a distorted absolute time percept so that participants perceived durations of staying in the experimental setup either as much longer or much shorter than in reality. This shows that a manipulation of body schema by visual delay strongly corresponds to a modulation of time percept, which was observed regardless of whether participants had been instructed in the experimental paradigm to stay a certain amount of time (90 s), or to stay as long as they wanted. Such an association between changes in body schema and time perception has previously been reported, for example in schizophrenic patients who display a "natural" deviation of body schema compared to healthy controls (Thakkar et al., 2011). Furthermore, time perception is affected in individuals whose body schema was modulated by means of meditation (Berkovich-Ohana et al., 2013). Time perception is also different after the consumption of drugs, which often also leads to changes in body schema (Edelmann and Terbuyken, 2002; Atakan et al., 2012). The results of the current study are thus in accord with previous concepts postulating an influence of body schema on time perception, but it is the first time that visual delay of one's own body has been used to manipulate time perception.

Note that while in the current study we manipulated the visual representation of the own body (thus an aspect of body schema), we more specifically also presented several visual representations of the own body at different time points simultaneously. This specific manipulation of body schema was thus also associated with feeding in references of the immediate past into working memory that is known to be crucial to time perception of durations in a seconds to minutes range (Buhusi and Meck, 2005; Allman et al., 2014). We argue that this has a parallel in the acoustic domain, where references to the immediate past through auditory-motor associations and operant conditioning play a major role in musical improvisations. Here a density of referential stimulation—different to the current visual experiment—would not be mediated by simultaneity (of several video signals), but by the mostly repetitive organization of music. Note, however, that no working memory related information was being 'actively' manipulated in the task, i.e., participants were for example not required to remember how their body

position was a few seconds ago, so the proposed link to working memory discussed here has yet to be considered rather weak.

Furthermore, the manipulation of own-body representations led to a number of other psychological outcomes that had previously been evoked with musical interventions (Rickard, 2004) or by drug administration (Parrott et al., 2007). Participants who were presented with their delayed body representations reported significantly higher perceived intoxication and arousal. These effects were (similarly to the observed effects on time perception) independent from participants being instructed to stay a certain amount of time (90 s), or stay as long as they wanted.

Additionally, the current study replicated the findings of previous research on human timing in the seconds to minutes range, which showed a general tendency to overestimate time in the seconds to minutes range (Wittmann et al., 2010). This tendency was found in all experimental conditions in the current experiment.

A limit of the current investigation is that effects of the time delays utilized are not systematically investigated. We just compared delay on/off. For future studies it would, for example, be interesting to identify which time delays work best for corrupting the perception of time. Further, we did not investigate if a general effect of non-delayed body representations compared to no visual feedback of one's body exists, another question that might be addressed in future studies. Previous research showed that attention can modulate perception of time (Brown, 1985). Accordingly, it is possible that attentional processes masked or contributed to effects by modulation of body schema on time perception, as we did not control for attentional load in this study. However, it seems rather unlikely that attention differed substantially between delay vs. non-delay groups because besides effects of group on time estimates, we also found effects of delay condition on outcomes completely unrelated to attentional processes, such as arousal or perceived intoxication. Previous research also reported a greater variability in time estimation tasks in individuals with an altered body schema, i.e., schizophrenic patients (Bolbecker et al., 2014). Due to the fact that we did only employ an in-between study design, we could not test for this directly in the current study. Nonetheless, descriptive differences in standard deviation between experimental groups may point to an effect of manipulation of body schema on time estimation performance variability. It would be interesting to focus more on this correlation of time perception and body schema in future work. In addition, it would be interesting to test the effects of delayed video representations over different durations in future studies, given that in the current study we only tested target durations of 90 s in the time awareness conditions.

In conclusion, we here report a novel method to modify the perception of time by modifying body schema through delayed video representations of the own body. In addition to effects on time perception, we furthermore report effects on perceived arousal and intoxication. We discuss possible parallels to the functioning of music in the acoustic domain with respect to presenting references to the immediate past.

---

[6]Bonferroni–Holm corrected alpha level: $\alpha' = 0.0125$.

Acknowledgments

References

Allman, M. J., Teki, S., Griffiths, T. D., and Meck, W. H. (2014). Properties of the internal clock: first- and second-order principles of subjective time. *Annu. Rev. Psychol.* 65, 743–71. doi: 10.1146/annurev-psych-010213-115117

Atakan, Z., Morrison, P. G., Bossong, M., Martin-Santos, R., and Crippa, J. A. (2012). The effect of cannabis on perception of time: a critical review. *Curr. Pharm. Des.* 18, 4915–4922. doi: 10.2174/138161211802884852

Barkley, R. A., Edwards, G., Laneri, M., Fletcher, K., and Metevia, L. (2001a). Executive functioning, temporal discounting, and sense of time in adolescents with attention deficit hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD). *J. Abnorm. Child Psych.* 26, 541–556. doi: 10.1023/A:1012233310098

Barkley, R. A., Murphy, K. R., and Bush, T. (2001b). Time perception and reproduction in young adults with attention deficit hyperactivity disorder. *Neuropsychology* 15, 351–360. doi: 10.1037//0894-4105.15.3.351

Bauermeister, J. J., Barkley, R. A., Martínez, J. V., Cumba, E., Ramírez, R. R., Reina, G., et al. (2005). Time estimation and performance on reproduction tasks in subtypes of children with attention deficit hyperactivity disorder. *J. Clin. Child Adolesc. Psychol.* 34, 151–162. doi: 10.1207/s15374424jccp 3401_14

Berkovich-Ohana, A., Dor-Ziderman, Y., Glicksohn, J., and Goldstein, A. (2013). Alterations in the sense of time, space, and body in the mindfulness-trained brain: a neurophenomenologically-guided MEG study. *Front. Psychol.* 4:912. doi: 10.3389/fpsyg.2013.00912

Berlucchi, G., and Aglioti, S. (1997). The body in the brain: neural bases of corporeal awareness. *Trends Neurosci.* 20, 560–564. doi: 10.1016/S0166-2236(97)01136-1133

Bernardi, N. F., Marino, B. F., Maravita, A., Castelnuovo, G., Tebano, R., and Bricolo, E. (2013). Grasping in wonderland: altering the visual size of the body recalibrates the body schema. *Exp. Brain Res.* 226, 585–594. doi: 10.1007/s00221-013-3467-3467

Bolbecker, A. R., Westfall, D. R., Howell, J. M., Lackner, R. J., Carroll, C. A., O'Donnell, B. F., et al. (2014). Increased timing variability in schizophrenia and bipolar disorder. *PLoS ONE* 9:e97964. doi: 10.1371/journal.pone. 0097964

Brown, S. W. (1985). Time perception and attention: the effects of prospective versus retrospective paradigms and task demands on perceived duration. *Percept. Psychophys.* 38, 115–124. doi: 10.3758/BF03198848

Buhusi, C. V., and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* 6, 755–765. doi: 10.1038/nrn1764

Edelmann, B., and Terbuyken, G. (2002). "Drogenkonsum als modifikation des körpererlebens," in *Wahrnehmen – Bewegen – Verändern: Beiträge zur Theorie und Praxis sport-, kórper- und Bewegungsorientierter Sozialer Arbeit*, eds G. A. Pilz and H. Böhmer (Hannover: Blumhardt), 159–193.

Glicksohn, J., and Hadad, Y. (2012). Sex differences in time production revisited. *J. Individ. Differ.* 33, 35–42. doi: 10.1027/1614-0001/a000059

Holmes, N. P., and Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cogn. Process.* 5, 94–105. doi: 10.1007/s10339-004-0013-3

Holmes, N. P., and Spence, C. (2006). "Beyond the body schema: visual, prosthetic, and technological contributions to bodily perception and awareness," in *Human Body Perception from the Inside Out,* eds G. T. Knoblich, M. M. Grosjean, and M. Shiffrar (Oxford: Oxford University Press), 15–62.

Holmes, N. P., and Spence, C. (2007). Dissociating body image and body schema with rubber hands. *Behav. Brain Sci.* 30, 211. doi: 10.1017/S0140525X07001501

Jackson, S. A., and Csikszentmihaly, M. (1999). *Flow in Sports: The Keys to Optimal Experiences and Performances*. Champaign, IL: Human Kinetics.

Kerns, K. A., McInerney, R. J., and Wilde, N. J. (2001). Time reproduction, working memory, and behavioral inhibition in children with ADHD. *Child Neuropsychol.* 7, 21–31. doi: 10.1076/chin.7.1.21.3149

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099. doi: 10.1126/science.1143439

O'Brien, E. H., Anastasio, P. A., and Bushman, B. J. (2011). Time crawls when you're not having fun: feeling entitled makes dull tasks drag on. *Pers. Soc. Psychol. Bull.* 37, 1287–1296. doi: 10.1177/0146167211408922

Parrott, A. C., Milani, R. M., Gouzoulis-Mayfrank, E., and Daumann, J. (2007). Cannabis and Ecstasy/MDMA (3,4-methylenedioxymethamphetamine): an analysis of their neuropsychobiological interactions in recreational users. *J. Neural Transm.* 114, 959–968. doi: 10.1007/s00702-007-0715-717

Pastor, M. A., Artieda, J., Jahanshahi, M., and Obeso, J. A. (1992). Time estimation and reproduction is abnormal in Parkinson's disease. *Brain* 115, 211–225. doi: 10.1093/brain/115.1.211

Pekala, R. J., Steinberg, J., and Kumar, V. K. (1986). Measurement of phenomological experience: phenomenology of consciousness inventory. *Percept. Mot. Skills* 63, 983–989. doi: 10.2466/pms.1986.63.2.983

Perbal, S., Droit-Volet, S., Isingrini, M., and Pouthas, V. (2002). Relationships between age-related changes in time estimation and age-related changes in processing speed, attention, and memory. *Aging Neuropsychol. Cogn.* 9, 201–216. doi: 10.1076/anec.9.3.201.9609

Peterburs, J., Nitsch, A. M., Miltner, W. H. R., and Straube, T. (2013). Impaired representation of time in schizophrenia is linked to positive symptoms and cognitive demand. *PLoS ONE* 8:e67615. doi: 10.1371/journal.pone.0067615

Rammsayer, T. H. (1997). On the relationship between personality and time estimation. *Pers. Indiv. Differ.* 23, 739–744. doi: 10.1016/S0191-8869(97)00117-117

Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychol. Music* 32, 371–388. doi: 10.1177/0305735604046096

Sheppard, K. E., and Parfitt, G. (2008). Patterning of physiological and affective responses during a graded exercise test in sedentary men and boys. *J. Exerc. Sci. Fit.* 6, 121–129.

Svebak, S., and Murgatroyd, S. (1985). Metamotivational dominance: a multimethod validation of reversal theory constructs. *J. Pers. Soc. Psychol.* 48, 107–116. doi: 10.1037//0022-3514.48.1.107

Thakkar, K. N., Nichols, H. S., McIntosh, L. G., and Park, S. (2011). Disturbances in body ownership in schizophrenia: evidence from the rubber hand illusion and case study of a spontaneous out-of-body experience. *PLoS ONE* 6:e27089. doi: 10.1371/journal.pone.0027089

Wearden, J. H., and Penton-Voak, I. S. (1995). Feeling the heat: body temperature and the rate of subjective time, revisited. *Q. J. Exp. Psychol. B* 48, 129–141. doi: 10.1080/14640749508401443

Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nat. Rev. Neurosci.* 14, 217–223. doi: 10.1038/nrn3452

Wittmann, M., Simmons, A. N., Aron, J. L., and Paulus, M. P. (2010). Accumulation of neural activity in the posterior insula encodes the passage of time. *Neuropsychologia* 48, 3110–3120. doi: 10.1016/j.neuropsychologia.2010.06.023

Older adults report moderately more detailed autobiographical memories

Robert S. Gardner[1,2], Matteo Mainetti[1] and Giorgio A. Ascoli[1,2,3]*

[1] Center for Neural Informatics, Structures, and Plasticity, Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA, [2] Psychology Department, George Mason University, Fairfax, VA, USA, [3] Molecular Neuroscience Department, George Mason University, Fairfax, VA, USA

Autobiographical memory (AM) is an essential component of the human mind. Although the amount and types of subjective detail (content) that compose AMs constitute important dimensions of recall, age-related changes in memory content are not well characterized. Previously, we introduced the Cue-Recalled Autobiographical Memory test (CRAM; see http://cramtest.info), an instrument that collects subjective reports of AM content, and applied it to college-aged subjects. CRAM elicits AMs using naturalistic word-cues. Subsequently, subjects date each cued AM to a life period and count the number of remembered details from specified categories (features), e.g., temporal detail, spatial detail, persons, objects, and emotions. The current work applies CRAM to a broad range of individuals (18–78 years old) to quantify the effects of age on AM content. Subject age showed a moderately positive effect on AM content: older compared with younger adults reported ~16% more details (~25 vs. ~21 in typical AMs). This age-related increase in memory content was similarly observed for remote and recent AMs, although content declined with the age of the event among all subjects. In general, the distribution of details across features was largely consistent among younger and older adults. However, certain types of details, i.e., those related to objects and sequences of events, contributed more to the age effect on content. Altogether, this work identifies a moderate age-related feature-specific alteration in the way life events are subjectively recalled, among an otherwise stable retrieval profile.

Keywords: autobiographical memory, memory content, aging, forgetting, recollection, episodic memory, word-cue technique

Introduction

Autobiographical memory (AM) refers to the recollection of personally experienced episodes specified in time, and has critical functions among adults of all ages (e.g., Pillemer, 1992; Bluck et al., 2005; Bluck and Alea, 2011; Waters, 2014). As any given AM is associated with a unique episode retrieved from a countless variety of experiences, any two AMs may differ greatly in terms of the type and amount of detail they contain (i.e., their content). For example, some memories contain a high degree of sensory and spatial information, whereas others comprise little sensory information but instead are associated with distinct thoughts and feelings; at the same time, certain memories are highly vivid and contain many details from numerous content categories, while others contain few and scattered elements.

The type and amount of subjective detail remembered in AM appear to be important dimensions of recall. For example, AM content is proposed to play an important role in source

monitoring (i.e., the process that determines the association between a memory and a particular context or source; Johnson and Raye, 1981; Johnson et al., 1988; Hashtroudi et al., 1990). As accurate source determinations may deteriorate in older adults (e.g., Cohen and Faulkner, 1989; Hashtroudi et al., 1990; Johnson et al., 1993; Gerlach et al., 2014), an understanding of age-related changes in AM content is essential. Nonetheless, quantitative characterization of AM content is lacking, most notably across a range of ages representative of the adult population, and across the life span of a given individual.

Autobiographical memory content has been probed in several ways. Some studies have assessed the veridicality of AMs by contrasting recalled details with verifiable event information. Resulting findings have highlighted the constructive and subjective nature of AM and suggest that the details retrieved by an individual may not match the event as it transpired in the material world (see e.g., Bartlett, 1932; Conway and Pleydell-Pearce, 2000). As such, many studies have focused on AM content as constructed and reported by the individual. Subjective memory content is typically measured by participant report using ordinal scales. For example, the Memory Characteristics Questionnaire (Johnson et al., 1988), Autobiographical Memory Questionnaire (Rubin et al., 2008), and Memory Experiences Questionnaire (Sutin and Robins, 2007) rate the perceived amount (or clarity) of sensory (e.g., visual, auditory), spatial, temporal, and emotional detail associated with a particular memory. Likewise, perceived event specificity has also been rated (e.g., Kopelman et al., 1989; Piolino et al., 2002; Sutin and Robins, 2007).

Relatively few studies have reported counts of the number of subjective details retrieved in AMs (Berntsen, 2002; Levine et al., 2002; St. Jacques and Levine, 2007; Addis et al., 2008, 2010); these studies typically use standard sets of cues (e.g., event-cues) to elicit memories, and subsequently collect from the participant written or spoken narratives of recalled experiences. Experimenters process each narrative (e.g., segment unique AMs), and score each memory for content. Several experiments using these techniques have shown that older (as compared with younger) subjects report fewer episodic details contained within memory for unique life experiences (e.g., Levine et al., 2002). However, AM content analyses have largely been confined to few and restricted life periods and age groups, and to memories that may not reflect naturalistic recall, e.g., those recalling experiences simulated in the laboratory (Hashtroudi et al., 1990) or those elicited using typical life event cues (Levine et al., 2002).

We previously introduced the Cue-Recalled Autobiographical Memory test (CRAM; see Gardner et al., 2012), in part, to address these limitations. CRAM elicits AMs using a modification of the word-cue technique (Crovitz and Schiffman, 1974). In contrast to traditional methods, word-cues are generated based on their usage frequency in spoken and written language in order to emulate naturalistic cues. Therefore, elicited AMs should be more closely matched to those recalled in everyday situations. Participants subsequently identify the age of each AM, and count the number of details recalled within specified features (e.g., temporal detail, spatial detail, persons, objects, emotions,

temporally linked events, and other contextual elements) similar to those used in previous designs (e.g., Johnson et al., 1988; Hashtroudi et al., 1990; Levine et al., 2002). Given CRAM's reliance on participants to specify what constitutes a detail within a feature category, this technique permits efficient data collection. This key advantage enables collection of larger data sets and relatively comprehensive AM coverage across the life span of numerous age groups. In addition, collecting count data (as opposed to relying on ordinal scales) facilitates interpretation of between-subjects and between-groups comparisons.

Despite its methodological differences (as compared with previous tests developed to probe AM), CRAM reliably reproduces several results of prior studies among young adults. For example, AMs cued by CRAM produce temporal distributions which completely replicate characteristics of those produced by traditional techniques, e.g., the retention interval and childhood amnesia (Rubin, 1982, 2000; also see Rubin et al., 1986; Rubin and Schulkind, 1997; Janssen et al., 2011). Moreover, AMs scored by CRAM show a temporal decay in content, a reported component of AM retrieval (e.g., Levine et al., 2002; Piolino et al., 2002; Janssen et al., 2011).

The current work builds on this prior research, which focused on college-aged subjects, by applying CRAM to individuals of various ages across the adult life span (18–78 years old). We utilized both in-person and Internet-based testing[1] to further enhance data collection. The resulting data provide numerical counts of AM content from a diverse subject pool that should expand our understanding of the relationship between aging and subjective recollection.

In particular, this research was conceived to describe age-related changes in subjective AM content, and to test the hypothesis that, on average, subjective reports of AM detail decrease both with the age of the event and with the age of the subject. Similar to studies that quantitatively described age-dependent modulation of the temporal distribution of AM retrieval (Rubin et al., 1986; Rubin and Wenzel, 1996; Rubin and Schulkind, 1997; Rubin, 2000) this work quantitatively characterizes age-dependent modulation of feature-specific recollection.

Application of CRAM to older subjects may also contribute to AM theory. The reminiscence bump is an increase in retrieval of AMs pertaining to episodes from adolescence to early adulthood and is most clearly observed in older adults (see Rubin et al., 1986, 1998; Jansari and Parkin, 1996; Janssen et al., 2005, 2011). Previous studies show that AMs from the bump, as collected using the word-cue technique, are not associated with enhanced phenomenological characteristics of recollection, e.g., vividness or re-living (Rubin and Schulkind, 1997; Janssen et al., 2011). However, whether content counts of these memories correlate with their retrieval probabilities remains an open question. For example, it is possible that memories rich with detail have relatively high association probabilities with a given memory cue, causing these AMs to be frequently accessed; that is, enhanced retrieval of bump memories may produce (or result from) enhanced content retrieval. More specifically, this research tests

---

[1]http://cramtest.info

the hypothesis that subjective content reported from bump AMs is increased compared with that reported from non-bump AMs. In addition, this approach and resulting data may be useful to inform theories of memory, e.g., multiple trace theory (Nadel et al., 2000).

Materials and Methods

The Cue-Recalled Autobiographical Memory Test

The Cue-Recalled Autobiographical Memory test is a computerized interactive test presented in web-browser format. It collects counts of the number of details (elements) within categories (features) that compose naturalistically word-cued AMs dated to specific life periods. Complete specifications of the test and instruction provided to participants have been previously reported[2] (for full detail see Gardner et al., 2012). Here, each section of the test is briefly described.

Prior to eliciting AMs, CRAM collects demographic information for each participant. Subjects are then presented the following definition of AM and subsequent instruction:

"Autobiographical memories are recollections of past episodes directly experienced by the subject. These memories should be of a brief, self-consistent episode of your life. An episode can be as short as a single snapshot and up to a few seconds long. . . . If the memory you think of refers to a typical and repeated episode that happened regularly or multiple times in your life, you can use it only if you can fixate on a specific individual event. If you can only recall the generic (repeated) event, look for another memory."

The test is designed to probe memory of unique personally experienced episodes, for example, memory for a conversation one had with friends at a recent dinner party, or memory of meeting one's spouse. This type of memory is considered episodic and has been frequently contrasted with memory for factual information about the world or oneself (semantic: Tulving, 1972, 1985).

Naturalistic word-cues are then presented to elicit memories. The participant reads through a list of seven words and labels the first recollection retrieved, for subsequent identification. Subjects are further instructed that the cued AM does not necessarily have to relate to any one word or to the entire list of words, but rather is "the first autobiographical memory that the words bring to mind." The list of word-cues is randomly selected from the British National Corpus[3], a compilation of 100 million written and spoken words (see Gardner et al., 2012 for word processing details). Thus, this procedure provides cues that are presented proportionally to word frequencies observed in everyday settings. For example, the word "waiting" appears in the corpus 474 times whereas the word "yard" appears 71 times. Therefore, "waiting" is ~6.7 times more likely than "yard" to be presented as a word cue. As the words are sampled randomly from the corpus each time a list is generated, the specific words selected may be

dramatically different for any two presentations within and across subjects.

Once AMs are cued and labeled, participants are presented with their AM labels one by one to date each memory. Specifically, the participant places each memory into one of 10 temporal bins, which segment his or her life span into 10 equal intervals (a similar binning procedure has been used previously: see McCormack, 1979; Howes and Katz, 1992; Gardner et al., 2012). If necessary, up to three temporal bins could be assigned to a single AM. To increase dating accuracy, the temporal range associated with each bin (i.e., the cutoff values computed from the subject's age) is presented in terms of time from the present, age of the participant, and month and year. However, as the size of a given temporal bin increases linearly with the participant's age at the time of testing (e.g., the size of a bin for a 20 year old is half that for a 40 year old), this procedure also reduces the temporal resolution associated with a memory in older (relative to younger) subjects.

The age of the participant at the time of a recalled event is estimated as the mid-point of its assigned temporal bin. The retrieval probability of an AM falling within each bin or age range is computed by dividing the number of AMs dated to each temporal period by the total number of dated AMs. Collectively, these measures are used to construct the temporal (life span) distribution of AMs. In this work, Recent AMs are defined as memories of events that occurred within the most recent 10 years of life; Remote AMs are defined as memories of events that occurred more than 10 years from the present moment. Overall, ~4% of AMs were assigned to more than one temporal bin (presumably resulting from a lack of confidence that the AM occurred within the temporal range associated with a single bin). This proportion mildly increased for Remote AMs (Remote: 5.0%; Recent: 3.7%, $p < 0.001$), suggesting a reduction in a given subject's confidence to date these older episodes. In addition, younger subjects dated AMs to multiple bins slightly more frequently (4.0%) than older subjects (3.0%; $p < 0.05$), potentially due to the age-dependent nature of a given bin's temporal interval size (the temporal range associated with a bin increases proportionally with the subject's age at the time of testing).

After AMs are dated, participants are once again presented with their AM labels to score the content associated with a selected memory. In particular, participants are instructed to count the number of details remembered within each of eight categories, i.e., *Things* (objects), *Feelings* (emotional details), *People* (unique individuals), *Places* (spatial details), *Times* (temporal details), *Episodes* (temporally linked events), *Contexts* (other contextual details), and *Details* (all remaining details, including actions). Participants provide counts of details rather than event descriptions. The order in which each category is presented is randomized for each person but fixed across all AMs for a given individual. The exact definitions of these eight categories were previously reported (Gardner et al., 2012) and are available at http://cramtest.info. In addition, CRAM provides, through clickable links, additional examples of what constitutes a detail within a given category and general scoring guidance. Every detail category is called a "feature," each reported detail

---

[2]http://cramtest.info

[3]http://www.natcorp.ox.ac.uk

associated with a given feature is referred to as an "element," and the summed number of elements across all features for a given memory is called "total content."

In-Person and Internet Testing

The Cue-Recalled Autobiographical Memory test was completed locally under experimenter supervision or remotely over the Internet[4]. From each subject who completed testing in person, 30 AMs were cued and dated, of which a subset of 10 was scored for content (see Gardner et al., 2012; the first two AMs were considered practice and not analyzed). AMs were selected for scoring to maximize life span coverage in the entire dataset. Specifically, a single AM was scored from each life span bin represented by a participant. From the participant's remaining AMs, memories were selected in order from the least to most represented bin (in the entire dataset across all participants) until 10 memories were scored in total; only for an initial sample of subjects ($n = 110$) the number of scored AMs was instead a function of the number of temporal bins represented.

In contrast to in-person testing, CRAM's online protocol offers subjects the choice of several test options (i.e., Atomic, Mini, Extended, and Full tests) which differ according to the number of AMs cued, dated, and scored. These options are included to promote test completion by suiting a wide range of subjects who may vary in their commitment and eagerness to participate. The Atomic test cues one AM which is dated and scored for content. At the end of the Atomic test, subjects are asked if they would like to complete the Mini or Full test. The Mini test cues five AMs, which are each dated and scored. At the end of the Mini test, subjects are invited to extend the test. If a subject agrees, an additional fifteen AMs are cued and dated, five of which are scored; this option is categorized as the Extended test (which in total cues and dates 20 AMs, and scores 10 for content). The Full test cues 20 AMs, each of which is dated (only nineteen subjects completed instead a different Full test version which cued and dated 30 AMs as outlined in the in-person protocol). Subsequently, a subset of 10 memories is scored for content. Memory selection for scoring in the Full and Extended tests follows the same rules as those for in-person testing. Given these selection rules, the proportion of scored AMs in a particular life span interval may differ from that typically retrieved. Thus, when presenting aggregate measures of content (i.e., those within a given subject group) not restricted to a particular life period, content values across dating bins are weighted according to the applicable AM temporal distribution.

During online data collection, participants are encouraged to complete the Full test (or Extended test, if opting initially for the Mini test). This is accomplished by pre- and post-test advertisement for the opportunity to explore an interactive summary report of one's results with the ability to make direct comparisons with results from specified age ranges, solely after completion of the Full (or Extended) test. We stress that, despite their variety (e.g., in duration), all test types provided subjects with the same instruction on AM classification, cueing, dating, and scoring (which were also identical to in-person testing).

When conducting analysis on an individual level, we assumed that a unique test ID corresponds to a unique user. By and large this assumption should be valid; however, it is likely that a subset of users took the test multiple times and were assigned distinct user IDs on each occasion. Moreover, undertaking the Mini or Full version of CRAM following completion of the Atomic format was not coded in the same way as extending the Mini test; in particular, those who opted for the Atomic test and subsequently decided to undergo longer testing were assigned a new user ID. We emphasize that those users who were interrupted or opted to take a break mid-test, however, maintained a single user ID. Likewise, users who opted to extend the Mini test were categorized as taking the Extended test and assigned one unique user ID. As data collected prior to and subsequent to the decision to extend the test was equivalent (data not shown) it appears that practice and in-depth knowledge of dating and scoring procedures did not influence retrieval as measured by CRAM. Unless indicated otherwise, data were collapsed across testing conditions and test types.

Participants

As CRAM is freely accessible online[5] and indexed by popular search engines, data are continuously collected from Internet-browsing individuals. To supplement these unsolicited data, additional individuals were actively recruited from the undergraduate population of George Mason University (GMU), from GMU staff and faculty, and from the local community, obtaining in all cases informed consent. With the exception of undergraduates, recruited subjects were given the choice to complete testing locally at GMU with a researcher present or remotely over the Internet. Recruited undergraduates (ages 18–36 years old) invariably completed the study for course credit and took the test under experimenter supervision; these data from recruited students have been reported previously (Gardner et al., 2012) and included here to best estimate AM content in relatively young subjects. However, the amount of data collected from this age range was substantially augmented by the current approach (exclusively through online testing) almost tripling the previous sample of scored AMs (when pooled together) from these younger subjects. No identifiable personal data were stored. All recruitment and testing procedures were approved by the GMU institutional review board.

In total, 17,482 AMs were dated from 2,561 unique test IDs (Mean Age = 34 years old, SD = 14, range: 18–78 years old; 67% female; 81% native English speakers). A subset of subjects ($n = 640$) dated AMs, but did not score their content, and were restricted to memory dating analysis. In addition, content measures from 560 Internet-collected AMs (spanning 192 unique test IDs) were inadvertently overwritten prior to back-up, and thus these AMs were also restricted to dating analysis (complete content measures from a subset of AMs from four of these 192 subjects were retained, however, and included in content analysis). After accounting for these events, a total of 6,492 AMs were scored for content (76% scored online) from 1,733 subjects. Fifty-two percent of Internet-scored AMs were collected from the

---

[4]http://cramtest.info

[5]http://cramtest.info

Full test, 20% from the Atomic test, 16% from the Mini test, and 12% from the Extended test; subject age was equivalent across test types ($p > 0.10$).

Data Screening

Data were further inspected to identify data entry errors or otherwise lazy and inauthentic reporting (e.g., see Gardner et al., 2012). Positive cases were removed from analysis. For example, AMs were excluded if a subject reported an identical number of elements for each of the eight features. In addition, all AMs were removed from seventeen participants whose scoring across the majority of their AMs reflected this pattern (201 AMs in total). Memories were also excluded from two participants who reported either 1 or 11 elements in each feature category across all scored AMs (14 AMs), and from two subjects who reported unique scores for each feature but identically scored all memories (15 AMs). Altogether, these exclusions totaled 232 scored AMs.

Subsequently, extreme total content values were identified as those greater than three times the Inter-Quartile-Range (IQR) above the 75th percentile within a given age range. Data meeting this criterion were considered outliers and excluded from analysis. This procedure was performed separately for AMs collected from each of the following age ranges: 18–25; 26–35; 36–45; 46–55; 56–65; 66–78 years of age. The outlier threshold ranged from 72 to 98 elements per memory depending on the age group (18–25 years old: 72 elements; 26–35 years old: 79 elements; 36–45 years old: 91 elements; 46–55 years old: 98 elements; 56–65 years old: 94 elements; 66–78 years old: 80 elements). This step resulted in the removal of 223 AMs: 78 AMs were excluded from 18 to 25 year old subjects (3.0% of the total within this age range), 52 AMs (3.5%) from 26 to 35 year old subjects, 39 AMs (4.0%) from 36 to 45 year old subjects, 33 AMs (5.1%) from 46 to 55 year old subjects, 16 AMs (4.5%) from 56 to 65 year old subjects, and 5 AMs (2.1%) from 66 to 78 year old subjects. **Table 1** provides a summary of the number of scored AMs ultimately retained for data analysis across test types and participant groups.

Statistics

Analyses were primarily conducted using memory as the unit of observation. Findings were corroborated using an individual subject level approach (see Results). Binary logistic regression was run to assess the effects of participant groups on AM retrieval probabilities across life periods (i.e., Recent and Remote). Bivariate regression was performed to evaluate correlation between measures of AM recall and participant age, and inter-feature relationships. ANOVA was performed to evaluate changes in subject demographics across test types and AM content across participant groups and life periods. Results were corroborated by using a generalized estimating equation approach (Davis, 2002); all conclusions were equivalent with those reported using a general linear model. Where applicable, for robustness analyses (i.e., analysis across genders, native languages, and testing conditions), subject age, and/or cuing and scoring order were assigned as covariates to control for variation in the outcome variable explained by these sources. Chi-square analysis with Yates correction was run to assess group differences in the proportion of AMs assigned to multiple life span bins. Cohen's d was calculated for each comparison to estimate effect size. Statistical significance was interpreted using the criterion of $p < 0.05$. False discovery rate correction was applied to multiple comparisons (Benjamini and Hochberg, 1995). Statistical analyses were performed using SPSS (IBM), Excel (Microsoft), and R (Dalgaard, 2008).

Results

AM Retrieval Probabilities are Modulated by Subject Age and Life Period

Autobiographical memory retrieval probabilities were analyzed against the age of the subject at the time of the recalled episode among various age groups (**Figure 1**). We observed a large proportion of AMs recalling recent events, which declined steeply with time from the present moment (retention interval). We also found a relative increase in the number of AMs dated to adolescence through young adulthood (the bump), and a relative absence of AMs from early childhood (childhood amnesia).

Moreover, these characteristics of the temporal distribution of AMs produced by CRAM changed with participant age. Retention of Recent AMs was strongest in younger subjects (18–25 years old) and decreased with increasing participant age ($p < 0.001$). For example, AMs dated to the most recent 10 years of life comprised ~77% of all AMs cued from 18 to 25 year old subjects, ~42% of those cued from 26 to 45 year

TABLE 1 | The distribution of scored autobiographical memories (AMs) collected across test types and participant groups.

	Younger subjects (18–45 years old)					Older subjects (46–78 years old)				
	In-person	Atomic	Mini	Extended	Full	In-person	Atomic	Mini	Extended	Full
n (Subjects)	188	699	122	58	203	6	211	47	21	69
Mean age	20.9	29.6	30.3	30.7	29.5	53.5	55.8	56.3	54.4	57.8
SD age	4.3	7.7	7.5	8.0	7.5	10.8	7.2	7.1	6.3	9.0
% Female	76	71	78	84	67	83	63	66	86	64
% Native	73	80	81	84	83	100	90	87	86	87
n (Memories)	1489	699	515	409	1747	45	211	201	147	574

The number of subjects and memories included in content analysis is shown as n. Percent Native indicates the percent of subjects who are native English speakers. SD, standard deviation.
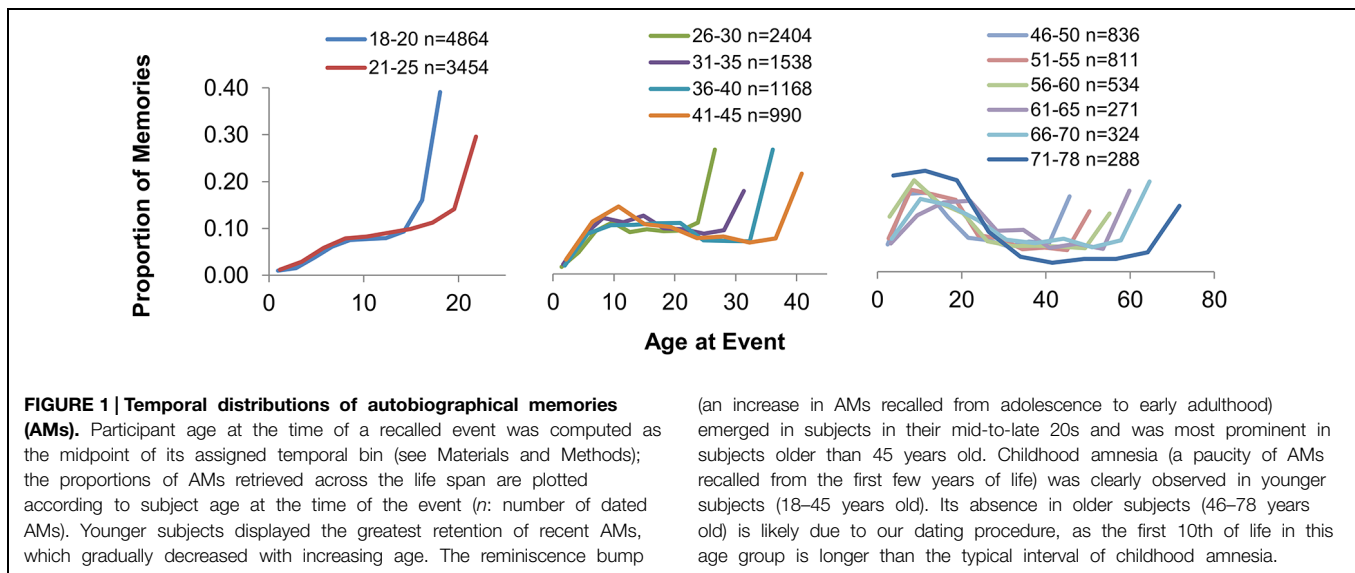
FIGURE 1 | Temporal distributions of autobiographical memories (AMs). Participant age at the time of a recalled event was computed as the midpoint of its assigned temporal bin (see Materials and Methods); the proportions of AMs retrieved across the life span are plotted according to subject age at the time of the event (*n*: number of dated AMs). Younger subjects displayed the greatest retention of recent AMs, which gradually decreased with increasing age. The reminiscence bump

(an increase in AMs recalled from adolescence to early adulthood) emerged in subjects in their mid-to-late 20s and was most prominent in subjects older than 45 years old. Childhood amnesia (a paucity of AMs recalled from the first few years of life) was clearly observed in younger subjects (18–45 years old). Its absence in older subjects (46–78 years old) is likely due to our dating procedure, as the first 10th of life in this age group is longer than the typical interval of childhood amnesia.

TABLE 2 | Autobiographical memory content and retrieval across life periods and age groups.

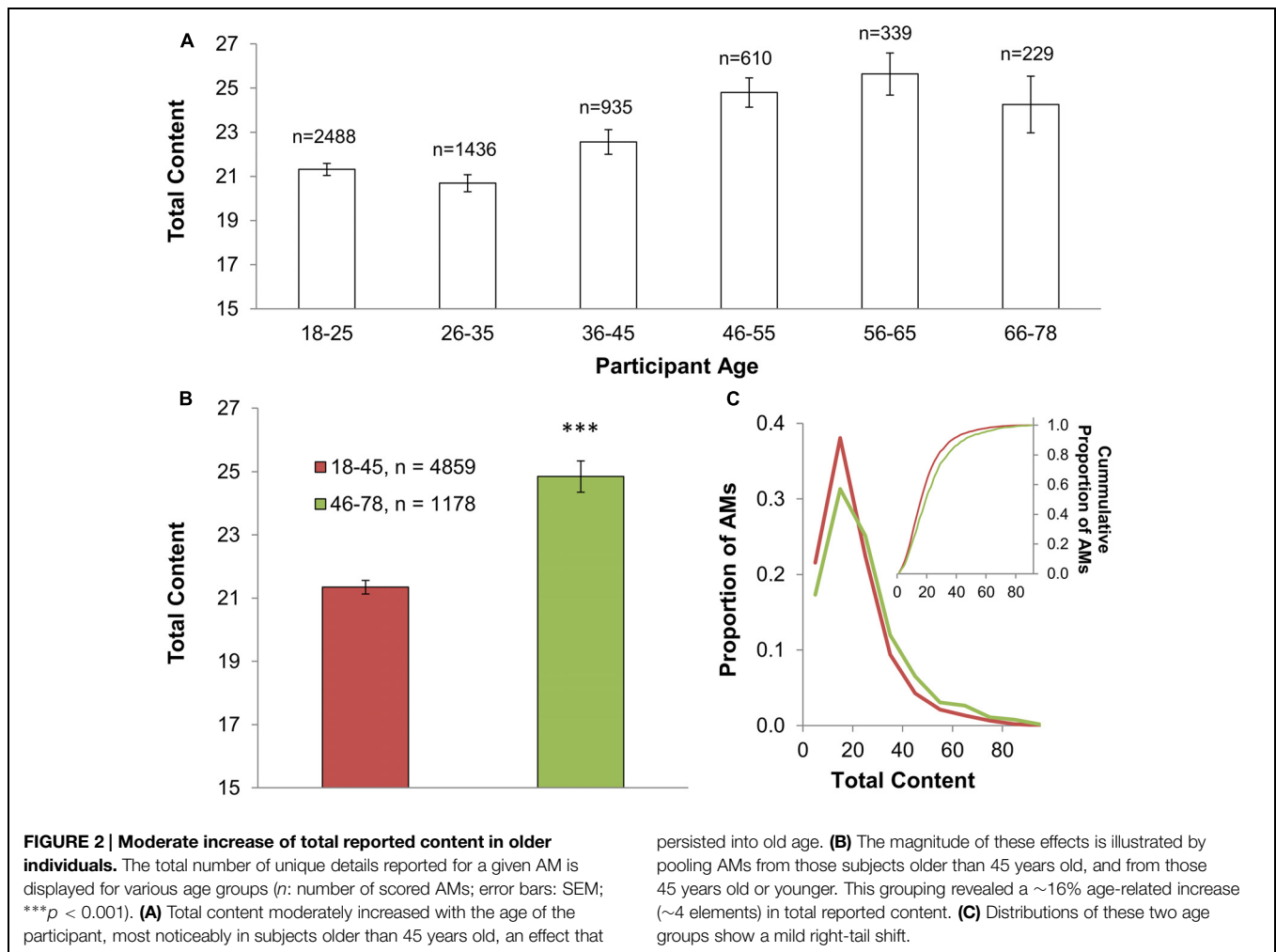|		Remote			Recent (10 years)		
		18–25 years old	26–45 years old	46–78 years old	18–25 years old	26–45 years old	46–78 years old
Total content	Mean	19.05	20.68	24.23	21.04	22.17	27.46
	SD	11.70	14.83	16.07	12.23	15.30	18.02
	CV	0.61	0.72	0.66	0.58	0.69	0.66
	Median	16	17	21	19	19	23
	IQR	13	17	18	15	16	23
AMs dated	*n* (%)	1890 (23%)	3549 (58%)	2469 (81%)	6428 (77%)	2551 (42%)	595 (19%)
AMs scored	*n*	826	1592	1016	1662	779	162

Recent AMs are those dated to the most recent 10 years; all remaining AMs are Remote (SD, standard deviation; CV, coefficient of variation; IQR, inter-quartile range; n, number of AMs within each subgroup).

old subjects, and ~19% of those cued from subjects older than 45 years (see **Table 2**). Furthermore, while absent in the two youngest age groups, the reminiscence bump emerged in subjects in their mid-to-late 20s, and was most evident in subjects older than 45 years of age (**Figure 1**). The peak of the bump corresponded to the years between ages 8 and 22, depending on the age group (**Figure 1**). Age ranges presented in each panel in **Figure 1** reflect the changing clarity of the reminiscence bump quantified as the ratio of the peak retrieval probability across life span bins (excluding those associated with the retention interval) to the subsequent minimal retrieval probability. On average, this ratio is undefined in 18–25 year olds (due to the lack of a minimum, reflecting the absence of a bump), greater than one in subjects 26–45 years old, and greater than two in adults older than 45. Childhood amnesia was observed in younger subjects (18–45 years old) as demonstrated by a notable drop in AMs dated to the first 10th of the life span (less than 2%). Its apparent absence in older subjects (i.e., 46–78 years old; **Figure 1**) is likely an artifact of our methodology. Specifically, the first tenth of life of an older individual extends beyond the relatively narrow temporal period associated with childhood amnesia, and thus limits our

ability to isolate and analyze memory for these very early life events.

Reported AM Content is Moderately Increased in Older Adults

A total of 6,037 scored AMs were analyzed for content (**Table 1**; see Section "Materials and Methods" for data screening procedures). On average, subjects reported ~22 elements (SD = 14) per AM. We found a mild yet significant positive relationship between total reported content and participant age ($r = 0.11$, $p < 0.001$). To further investigate this finding, memories were separated into six age groups (see **Figure 2A**). Upon comparison to the youngest age group (18–25 years old: Mean Total Content = 21.3, SD = 12.1), adults older than 45 years old reported a greater number of remembered details (46–55 years old: $M = 24.8$, SD = 16.3, $d = 0.24$; 56–65 years old: $M = 25.6$, SD = 16.1, $d = 0.30$; 66–78 years old: $M = 24.3$, SD = 16.8, $d = 0.20$); in contrast, reports of content from individuals younger than 46 years old (26–35 years old: $M = 20.7$, SD = 14.0; 36–45 years old: $M = 22.6$, SD = 16.4) were found to be comparable to those collected from the youngest group ($d = 0.05$, and $d = 0.09$, respectively).

FIGURE 2 | Moderate increase of total reported content in older individuals. The total number of unique details reported for a given AM is displayed for various age groups (n: number of scored AMs; error bars: SEM; ***p < 0.001). **(A)** Total content moderately increased with the age of the participant, most noticeably in subjects older than 45 years old, an effect that persisted into old age. **(B)** The magnitude of these effects is illustrated by pooling AMs from those subjects older than 45 years old, and from those 45 years old or younger. This grouping revealed a ∼16% age-related increase (∼4 elements) in total reported content. **(C)** Distributions of these two age groups show a mild right-tail shift.
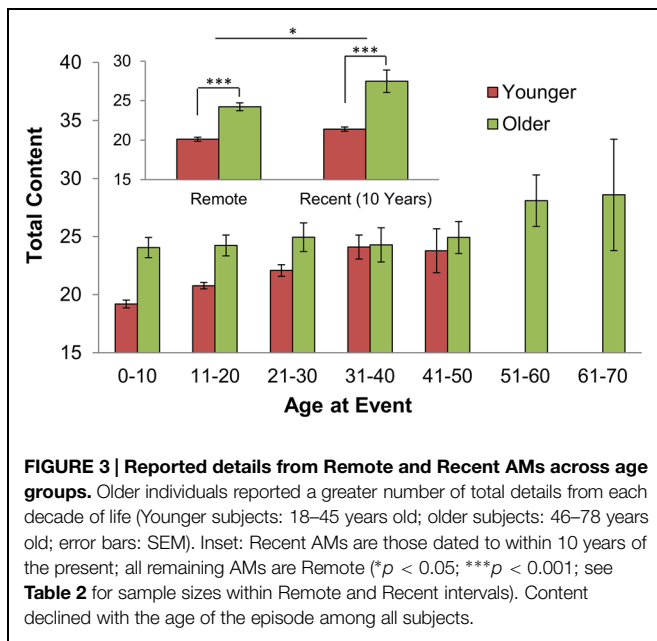
Collectively, these findings suggest a non-linear effect of age on total reported content; the age-related increase in the amount of detail reported in typical autobiographical recollection emerges most noticeably in subjects in their mid-to-late 40s and persists into old age (i.e., late 70s; **Figure 2A**). Given this pattern, to simply describe the magnitude of these effects, AMs were divided into two groups: those collected from subjects 45 years old or younger (Mean Age = 27, SD = 8, range: 18–45 years old; 72% female, 80% native English speakers) and those from subjects 46 years old or older (Mean Age = 57, SD = 8, range: 46–78 years old; 65% female, 89% native English speakers; **Figure 2B**). This grouping revealed a ∼16% increase in the number of reported details for a given AM in older compared with younger subjects (∼25 vs. ∼21; $p < 0.001$, $d = 0.23$; **Figure 2B**; **Table 2**). While reported content was quite variable from memory to memory, the coefficient of variation was similar across all ages (∼0.6–0.7; **Table 2**). The distributions of total content underscore the moderate effect of age on reported details (**Figure 2C**). Half of all scored memories in younger subjects were comprised of ∼16 or fewer elements; half of all scored AMs in older subjects contained more than ∼20 elements.

Total Content Decays with the Age of the Memory Across Younger and Older Subjects Alike

Older subjects reported significantly more content than younger subjects for memories of all life periods, both when comparing equivalent decades of life (**Figure 3**), and when comparing relative life periods (e.g., Recent: the most recent 10 years, and Remote: > 10 years from the present; see Materials and Methods; **Figure 3** Inset; **Table 2**). The numerical results of content analysis of Recent and Remote AMs only marginally fluctuated depending on how these temporal intervals were defined (e.g., restricting Recent AMs to the most recent 5 years; restricting Remote AMs to the first decade of life). In addition, all reported conclusions remained unchanged and did not depend on the particular grouping applied (data not shown).

Older subjects reported ∼24 elements from AMs dated to the first two decades of life compared with ∼20 elements in younger subjects ($p < 0.001$; $d = 0.27$). Likewise, Recent AMs from older subjects were comprised of ∼27 elements compared with ∼21 elements in those from younger individuals (**Figure 3** Inset; $p < 0.001$; $d = 0.38$). Total content declined with the age of the episode among all age groups. For example, Remote AMs were

FIGURE 3 | Reported details from Remote and Recent AMs across age groups. Older individuals reported a greater number of total details from each decade of life (Younger subjects: 18–45 years old; older subjects: 46–78 years old; error bars: SEM). Inset: Recent AMs are those dated to within 10 years of the present; all remaining AMs are Remote (*$p < 0.05$; ***$p < 0.001$; see **Table 2** for sample sizes within Remote and Recent intervals). Content declined with the age of the episode among all subjects.

comprised of significantly fewer details compared with Recent AMs ($p < 0.01$; **Table 2**; **Figure 3**). These results confirm those previously reported in college-aged subjects (Gardner et al., 2012) and extend these findings to older adults.

Features Selectively Contribute to the Age-Related Increase in Total Content

A relatively high proportion of AM content (~46%; ~10.1 elements) was associated with *Places*, *Things*, and *People*. In contrast, *Times*, *Contexts*, and *Episodes* were less represented, together comprising just ~29% (~6.4 elements). *Details* and *Feelings* were close to the average at ~13% and ~12%, respectively (**Figure 4A**). Largely in line with these distributions, the proportion of AMs containing at least one element of a particular feature was highest for *Places* (98%) followed by *People* and *Feelings* (92% each), *Things* (89%), *Details* (84%), *Times* (83%), *Contexts* (78%), *and Episodes* (63%). Feature variation was higher than that observed for total content, but equivalent across age groups (*Mean Feature CV* = 1.0). All these findings uphold those previously reported in younger subjects (Gardner et al., 2012).

Adding to this research, we found that older adults reported a greater number of elements among all features. *Episodes* and *Things* showed the most prominent age-related content increase (~30% and 27%, respectively; $p < 0.001$), while content associated with *People* and *Times* remained close to that observed in young subjects ($p > 0.10$; **Figures 4A,B**). These data collectively indicate that the age-related increase in total content is not uniformly distributed across features (in terms of either absolute or relative value). In particular, *Things* (24%), *Episodes* (16%), and *Details* (15%) contributed more substantially to the age-related increase observed in total content (**Figure 4C**), whereas contributions from *People* (4%)
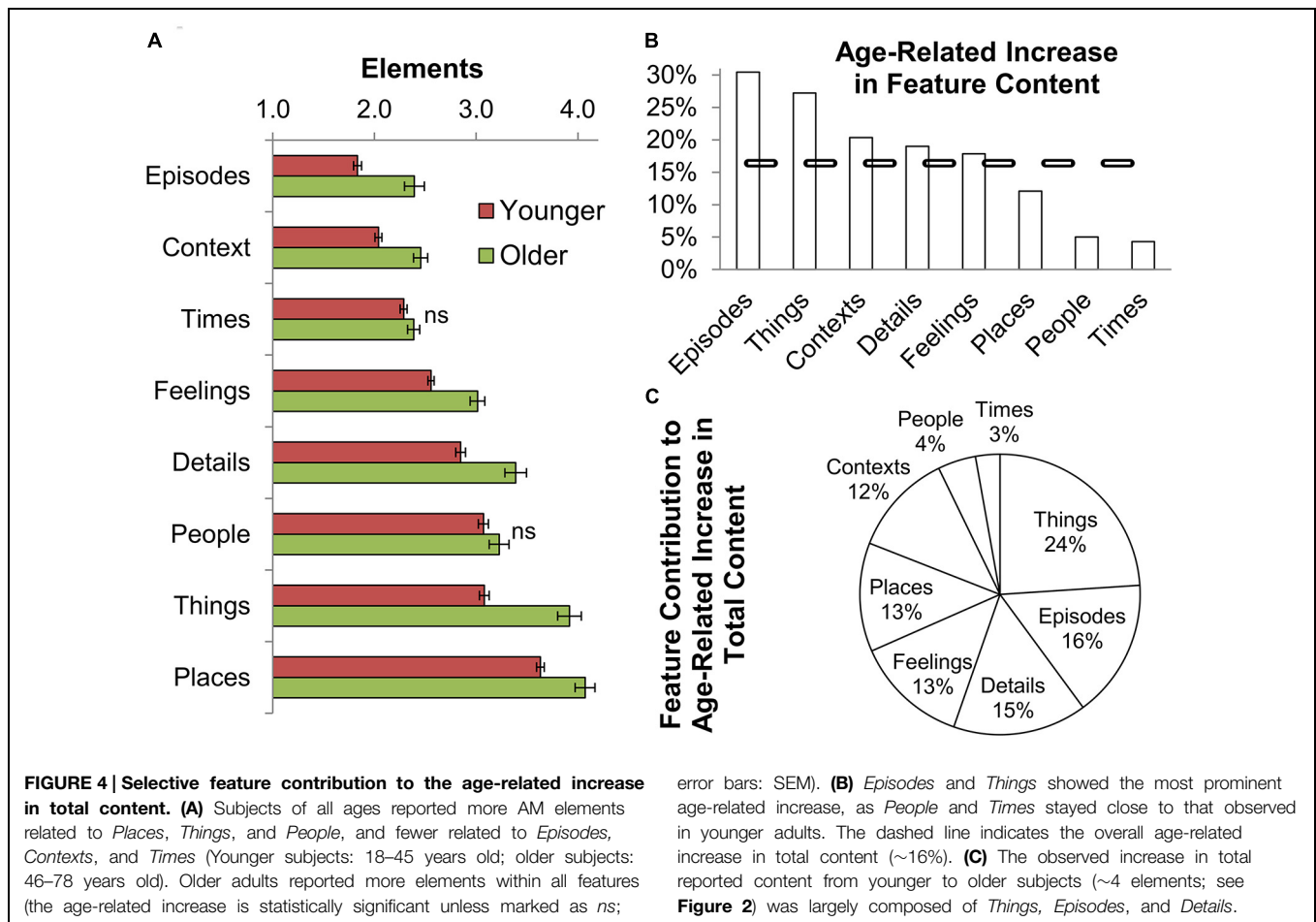
and *Times* (3%) were definitively smaller. The proportion of AMs containing at least one element from a given feature was significantly higher in older compared with younger subjects for *Feelings, Things, Details, Contexts,* and *Episodes* (~5% higher on average; $p < 0.001$) but not for *People, Places*, and *Times* ($p > 0.10$). These feature distributions were consistent between Remote and Recent AMs in both younger and older subjects (Supplementary Figure S1); notably, the feature *Times* appeared to be least resilient to temporal decay of content among all subjects.

People is a Relatively Independent Feature of Recall Among All Ages

Content correlation analysis showed positive relationships among all features (*mean Pearson r* = 0.37). Moreover, these relationships were similar across age groups (younger: $r = 0.36$; older: $r = 0.40$) as well as between Remote and Recent intervals (Remote: $r = 0.39$; Recent: $r = 0.37$). To further evaluate feature dependence, the correlation between each feature and all other content was computed. The average of these values across all features was equivalent between age groups (younger: $r = 0.53$; older: $r = 0.57$; Supplementary Figure S2A) and across life periods (not shown). However, the feature *People* exhibited a comparatively mild relationship ($r = 0.37$; 32% less than the average: Supplementary Figure S2B) across all conditions, suggesting that it is a relatively independent component of recall. The inter-feature relationships found here confirm those previously reported among younger subjects (Gardner et al., 2012) and extend these findings to individuals distributed across the life span.

Estimates of Retrieved Content across Age Groups and Life Periods

This work provides numerical description of AM retrieval probabilities and reported content associated with distinct life periods. Combining these two measurements, we can estimate the relative distribution of retrieved content across temporal intervals and age groups (**Figure 5**). Specifically, given the number of elements typically reported for a single retrieved AM, for each of the various age groups (as outlined in **Figure 2A** and plotted in **Figure 5** by mean age), we computed a probabilistic content distribution among life periods. For example, when experiencing AM, a typical 70 year old, on average, reports a similar content amount from his or her middle teenage years compared with that from the last few years of his or her life. In contrast, content from events dated to the most recent 2 years of the life of a 21 year old is ~5 times more represented in memory than that associated with events from his or her middle teenage years. Assuming that the frequency of AM recollection is stable across age groups (e.g., see Gardner and Ascoli, 2015), we can further estimate the relative probability that a particular recalled element is associated with a certain age group and life period. For example, the likelihood for a 30 year old to retrieve content from his or her early 20s (~1.3%) is equivalent to that for a 60 year old to retrieve content from his or her early thirties.

FIGURE 4 | Selective feature contribution to the age-related increase in total content. (A) Subjects of all ages reported more AM elements related to *Places*, *Things*, and *People*, and fewer related to *Episodes*, *Contexts*, and *Times* (Younger subjects: 18–45 years old; older subjects: 46–78 years old). Older adults reported more elements within all features (the age-related increase is statistically significant unless marked as *ns*; error bars: SEM). **(B)** *Episodes* and *Things* showed the most prominent age-related increase, as *People* and *Times* stayed close to that observed in younger adults. The dashed line indicates the overall age-related increase in total content (~16%). **(C)** The observed increase in total reported content from younger to older subjects (~4 elements; see **Figure 2**) was largely composed of *Things, Episodes*, and *Details*.

The Bump is Unrelated to Changes in Total Content and Feature Content

Autobiographical memories found within the reminiscence bump may have distinct recall characteristics. For example, AMs that compose the bump may be comparatively rich with recalled detail. Such a finding would explain, at least in part, why this life period plays a particularly prominent role in subjective experience. We evaluated reported content from AMs within and beyond life periods associated with the bump in older subjects (46–78 years old) for whom this phenomenon was most pronounced. In this age range, we observed a relatively high AM retrieval probability (excluding temporal bins associated with the retention interval) from ages 11–20 and a relatively low probability from ages 31–40 (see **Figure 1**). However, measures of total content from these life periods did not significantly differ ($M = 24.90$ and $M = 25.28$, respectively; $p > 0.10$, $d = 0.02$; **Figure 3**). The composition of memories dated to these life periods was also quite stable ($p > 0.10$), with any given feature showing a deviation of ~2% or less (*Mean deviation*: ~1%).

All the main findings of this work are robust to gender, native language, and changes in experimental procedures (see Supplementary Tables S1 and S2). For example, total reported content moderately increased with age both for females and males, as well as for native and non-native English speakers (see

Supplementary Material). However, several minor quantitative distinctions in AM recall were observed between males and females, between native and non-native English speakers, and among testing conditions. A full review of these results is included as Supplementary Material.

Age Effects are Corroborated Using Subject As the Level of Analysis

As prior reports suggest that different individuals have distinct recollection experiences (Rubin et al., 2004), the main conclusions of this work were evaluated using a subject-level analysis. Given the relatively large amount of variability in content scores from memory to memory (see **Table 2**), the analysis was restricted to those subjects who scored five or more memories (18–45 years old: $n = 496$; 46–78 years old: $n = 118$). This restriction permitted measures that more closely represent typical recollection from a given individual. In addition, for inclusion in comparisons across Recent and Remote intervals, subjects were required to have an AM scored in each temporal period (18–45 years old: $n = 420$; 46–78 years old: $n = 69$). Content measures reported independent of temporal periods were weighted according to the temporal distribution of AM retrieval computed separately for each subject on the basis of his or her dated AMs.
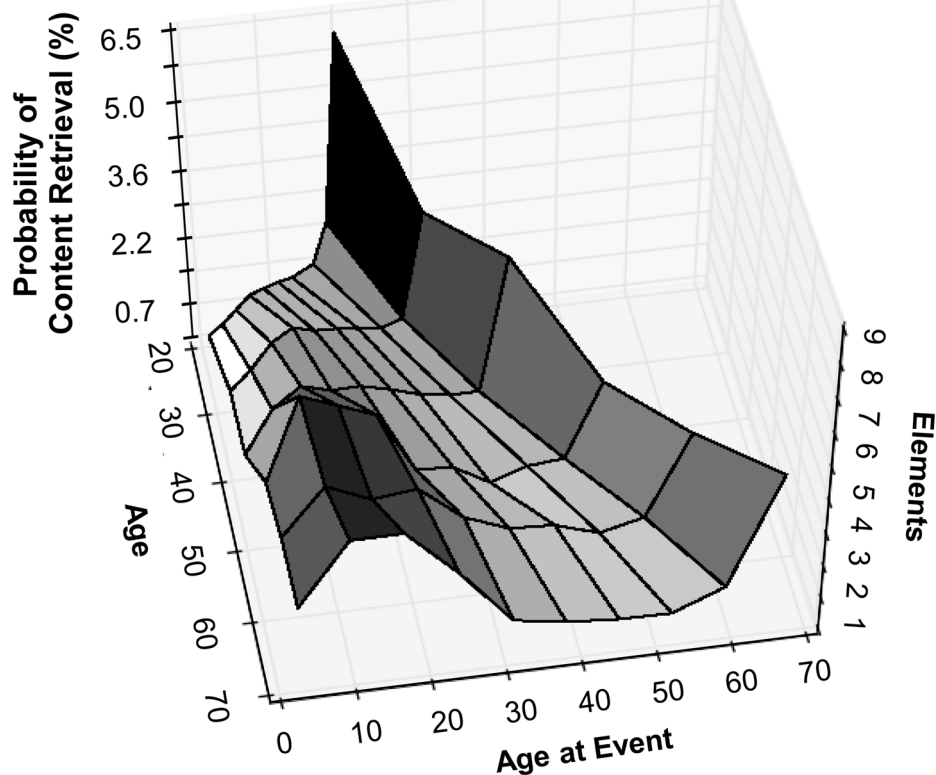
FIGURE 5 | Estimated distribution of retrieved content across age groups and life periods. Content measures within each temporal bin for each age group were normalized to their applicable retrieval probability. In particular, the number of elements per life period was computed for each of the six age

ranges outlined in **Figure 2A** (and plotted here by mean age within each group). These same computed values (normalized to all data points) provide estimates of the likelihood that any given amount of retrieved content is associated with a particular age and life period (shown as content retrieval probabilities).

Age effects on the temporal distribution of AMs and AM content are confirmed by subject-level analysis. In particular, the proportion of AMs from Recent life intervals was relatively high in young subjects (18–25 years old: ∼77%) and decreased with increasing age (26–45 years old: ∼40%; 46–78 years old: ∼17%). Moreover, older adults (46–78 years old) reported ∼15% more elements than younger (18–45 years old) subjects (Mean ± SD: 25 ± 13 compared with 21 ± 10 details; $p < 0.01$), replicating age differences found using memory as the observational unit. Remote memories contained fewer elements (18 ± 11) than Recent memories (23 ± 13; $p < 0.001$) across all subjects, and older compared with younger subjects reported more elements from Recent (Older subjects: 29 ± 18; Younger subjects: 22 ± 11; $p < 0.001$) and Remote events (Older subjects: 23 ± 13; Younger subjects: 18 ± 10; $p < 0.001$). In addition, an age-related increase in total reported content was most drastic for Episodes (∼31%) and Things (∼30%), whereas considerably less dramatic for Times (< 1%) and People (∼4%), closely resembling findings using memory as the unit of analysis. Likewise, total content scores from older subjects did not discriminate between bump and non-bump memories (e.g., comparing content from AMs dated to when subjects were 11–20 years old with those dated to when subjects were 31–40 years old; $p > 0.10$; data not shown).

Discussion

This work provides quantitative measures of reported AM content from individuals that represent a substantial segment of the adult population (18–78 years old). This was accomplished using CRAM, an instrument designed to collect counts of details that fall within specified features from naturalistically elicited AMs dated to particular life periods. Relying on participant counts of memory content (rather than experimenter-scored participant narratives) facilitates data collection and thus enables fine-scale analysis of AMs (e.g., as shown in **Figure 5**).

We previously demonstrated that CRAM replicates several noted observations of AM recall (e.g., on the retention of recent AMs and their associated content; Rubin and Schulkind, 1997; Levine et al., 2002; Piolino et al., 2002; Janssen et al., 2011). The present study upholds our previous findings from college-aged subjects. In particular, the current estimate of AM total content (∼21 elements) among younger subjects is almost identical to that reported (∼20 elements) by Gardner et al. (2012). In addition, the current work supports previous findings on the temporal decay of content (Remote AMs contain fewer elements), AM content variability (from memory to memory, total content remains relatively stable compared to feature content), AM

composition (*Places, Things, and People* are prominent features of recall), and inter-feature correlations (*People* is a relatively independent recall characteristic).

Extending CRAM to older adults, we replicated prior reports of age-modulation of the temporal distribution of AMs. Specifically, Recent AMs were considerably less likely to be recalled in older subjects (Rubin and Schulkind, 1997; Janssen et al., 2011), and the reminiscence bump emerged in subjects in their mid-to-late 20s. While some studies have found that the bump is not apparent until ∼40 years of age, using relatively small temporal bins, Janssen et al. (2011) reported a similar onset to that found here. In addition, the temporal interval associated with the bump using CRAM (i.e., 8–20 years old) is consistent with these studies.

Total content moderately but significantly increased with subject age. Older adults reported ∼25 details for a given AM, ∼4 elements more than the number reported from younger subjects. Moreover, this age-related increase in content was observed for Recent and Remote memories, and was most drastic for the features *Episodes* (sequences of event) and *Things* (objects) while negligible for *People* (unique individuals) and *Times* (temporal detail). Altogether, these data quantitatively describe an age-associated shift in the reported details of subjectively remembered events.

As older adults reported more content than younger adults from memories that originated in the same decade of life (i.e., those AMs that have similar ages of encoding), these findings appear to highlight the re-constructive nature of AM (Bartlett, 1932; Conway and Pleydell-Pearce, 2000; Hupbach et al., 2007). This interpretation assumes that the initial number of encoded event details and encoding depth are similar between age groups. However, as with all cross-sectional aging research, any inter-group differences may reflect generational differences rather than (or in addition to) changes that occur among individuals across their life span.

It also remains possible that older and younger adults used divergent strategies to establish feature counts. However, the finding that the age effects on content were feature-specific (see **Figure 4**) argues against a general change in interpretation of CRAM's instruction, and/or adjustment in content evaluation (e.g., a pervasive tendency for older individuals to report higher scores). Similarly, older subjects may be more motivated to recall and report event details. Nevertheless, the selection of online test types (which differed in their time commitment) was equivalent across age groups, suggesting that motivation may not underlie our findings. A similar conclusion can be drawn from the finding that content data collected in-person were equivalent to those collected online (see Supplementary Material), assuming that, on average, testing conditions correlate with task motivation. Further work, including the use of longitudinal designs, is required to clarify the mechanisms underlying the current findings.

Older individuals are proposed to have altered narrative attention and to tell more interesting life stories (James et al., 1998). Although more detail is not always better, inclusion of information about sequences of happenings (the feature that was most strongly augmented from younger to older subjects)

within a life narrative enables placement of an episodic snapshot into a broader context of surrounding events (and may enhance storytelling). It would be interesting to establish how our findings on feature-specific age-related modulation of reported AM content compare with the types and amount of detail shared through social communication of event memories and during the narration of life stories.

Despite the moderate change in total reported content, several properties of recollection were stable across age groups. In particular, AMs from younger and older subjects, and those from Remote and Recent life periods, showed similar feature distributions. In addition, among all ages, fewer details were reported from Remote than from Recent AMs. Thus, these data are indicative of two independent age effects on reported AM content: a positive correlation with the age of the individual and a negative one with the age of the event. These findings confirm and extend prior studies (Janssen et al., 2011; Gardner et al., 2012).

Additionally, independent of age, almost all AMs were reported to have at least one detail related to location, and nine out of 10 memories were reported to include some information about people, objects, and feelings, suggesting that these features are quite remarkable and/or at the core of subjective recall. In contrast, less than two-thirds of memories were reported to include sequential events. The feature *People* was also relatively independent, further demonstrating its unique role in memory; among the proposed functions of AM, remembrance of the individuals associated with specific life experiences is essential to form and maintain social relationships (see Pillemer, 1992; Bluck et al., 2005; Bluck and Alea, 2011; Waters, 2014).

Combining the observed temporal distributions of AM retrieval with measure of reported AM content permits computation of detailed probability estimates of reporting an element from a given life period at a particular age (**Figure 5**). For instance, this approach quantifies how likely it is for a detail retrieved by a 60 year old to be associated with an episode from his or her 20s (∼1.7%). Moreover, we can address questions on how these probability distributions change with subject age. For instance, how does the previously computed probability compare with the likelihood that the same amount of content retrieved by a 20 year old stems from a relatively recent event?

Several factors are proposed to account for the high accessibility of memories that compose the reminiscence bump, e.g., neurocognitive development, cultural influence, and life span changes in encoding efficiency (Rubin et al., 1998; Schrauf and Rubin, 2000; Berntsen and Rubin, 2004; Rathbone et al., 2008; Bohn and Berntsen, 2010; Janssen et al., 2015). We add to an understanding of the bump by showing that neither reported counts of detail across all features nor counts within individual features explain or are explained by the relatively high probability of recollection associated with this life period. These data are in line with previous studies that have demonstrated that AMs within the bump do not have higher ratings of certain characteristics of recollection (e.g., vividness, rehearsal, reliving, novelty, emotionality; Rubin and Schulkind, 1997; Janssen et al., 2011).

Past approaches reporting counts of AM content have predominantly focused on the distinction between episodic and

semantic retrieval (Levine et al., 2002; Piolino et al., 2002; Addis et al., 2008, 2010; also see Tulving, 1972, 1985). Episodic memory recounts a unique personally experienced event, with some form of contextual information (e.g., spatiotemporal detail). Semantic memory recalls abstracted knowledge of the world or of oneself (generally acquired from repeated experiences) that does not describe or call to mind a unique episode. This distinction is highlighted by case reports of neurocognitive deficits following targeted brain damage (Rosenbaum et al., 2005) and is present in numerous theories of AM (e.g., Conway and Pleydell-Pearce, 2000). Using a narrative scoring technique of event-cued AMs specified to life periods, Levine et al. (2002) found that, compared to younger subjects, older adults report fewer episodic but a similar or greater number of semantic details from typical memories. This age effect on episodic detail appeared to be feature-dependent as it was absent (but never reversed) for some features (e.g., *Times*). Contrasting accounts, however, have been reported. Hashtroudi et al. (1990) found that older adults reported more content for feelings and thoughts (albeit these same individuals showed a reduction in sensory-perceptual recollection). In addition, Janssen et al. (2011) found that ratings of AM vividness and re-living, proposed indices of episodic remembering were higher in older subjects (also see Rubin and Schulkind, 1997; Rubin and Berntsen, 2009). Direct comparison of memory content scores obtained using a variety of quantitative and qualitative approaches will be a useful endeavor to reconcile the apparently discrepant findings in AM recollection across age groups.

We emphasize that CRAM was broadly designed to measure the details that a participant considers part of an AM, i.e., the subjective content associated with a remembered life event. As such, CRAM does not classify reported elements as episodic or semantic. However, each feature definition was worded to collect the details that compose a memory for a temporally specific event; likewise, the general guidance provided to subjects emphasized reporting of detail unique to the specified episode (see Materials and Methods; Gardner et al., 2012). As CRAM collects reports of retrieved subjective detail, this approach also contrasts with those that aim to collect "true" or verifiable detail or those which collect all potentially retrievable details associated with an event (e.g., Mello and Fisher, 1996; Levine et al., 2002). We further stress that as counts of memory details are provided by the subject, although the results are in line with several prior findings, content measures as reported here may systematically differ from the actual numbers of details that are successfully retrieved from a given event memory. As CRAM's instruction was identical between all subjects, however, relative measures between and within age groups should reflect genuine changes in subjective memory.

Altogether, the data presented here provide previously inaccessible fine-scale quantitative characterizations of the reported subjective content of AMs as a function of the age of an individual and the age of a memory. These characterizations point to a moderate but significant age-associated feature-specific shift in how one's life story is perceived and recounted.

Acknowledgments

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00631/abstract

References

Addis, D. R., Musicaro, R., Pan, L., and Schacter, D. L. (2010). Episodic simulation of past and future events in older adults: evidence from an experimental recombination task. *Psychol. Aging* 25, 369–376. doi: 10.1037/a0017280

Addis, D. R., Wong, A. T., and Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychol. Sci.* 19, 33–41. doi: 10.1111/j.1467-9280.2008.02043.x

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology.* Cambridge: Cambridge University Press.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B Methodol.* 57, 289–300. doi: 10.2307/2346101

Berntsen, D. (2002). Tunnel memories for autobiographical events: central details are remembered more frequently from shocking than from happy experiences. *Mem. Cognit.* 30, 1010–1020. doi: 10.3758/BF03194319

Berntsen, D., and Rubin, D. C. (2004). Cultural life scripts structure recall from autobiographical memory. *Mem. Cognit.* 32, 427–442. doi: 10.3758/BF03195836

Bluck, S., and Alea, N. (2011). Crafting the TALE: construction of a measure to assess the functions of autobiographical remembering. *Memory* 19, 470–486. doi: 10.1080/09658211.2011.590500

Bluck, S., Alea, N., Habermas, T., and Rubin, D. C. (2005). A TALE of three functions: the self? Reported uses of autobiographical memory. *Soc. Cogn.* 23, 91–117. doi: 10.1521/soco.23.1.91.59198

Bohn, A., and Berntsen, D. (2010). The reminiscence bump reconsidered: children's prospective life stories show a bump in young adulthood. *Psychol. Sci.* 22, 197–202. doi: 10.1177/0956797610395394

Cohen, G., and Faulkner, D. (1989). Age differences in source forgetting: effects on reality monitoring and on eyewitness testimony. *Psychol. Aging* 4, 10–17. doi: 10.1037/0882-7974.4.1.10

Conway, M. A., and Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychol. Rev.* 107, 261–288. doi: 10.1037/0033-295X.107.2.261

Crovitz, H. F., and Schiffman, H. (1974). Frequency of episodic memories as a function of their age. *Bull. Psychon. Soc.* 4, 517–518. doi: 10.3758/BF03334277

Dalgaard, P. (2008). *Introductory Statistics with R,* 2nd Edn. New York, NY: Springer.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements.* New York, NY: Springer Science & Business Media.

Gardner, R. S., and Ascoli, G. A. (2015). The natural frequency of human prospective memory increases with age. *Psychol Aging* doi: 10.1037/a0038876 [Epub ahead of print].

Gardner, R. S., Vogel, A. T., Mainetti, M., and Ascoli, G. A. (2012). Quantitative measurements of autobiographical memory content. *PLoS ONE* 7:e44809. doi: 10.1371/journal.pone.0044809

Gerlach, K. D., Dornblaser, D. W., and Schacter, D. L. (2014). Adaptive constructive processes and memory accuracy: consequences of counterfactual simulations in young and older adults. *Memory* 22, 145–162. doi: 10.1080/09658211.2013.779381

Hashtroudi, S., Johnson, M. K., and Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychol. Aging* 5, 119–126. doi: 10.1037/0882-7974.5.1.119

Howes, J. L., and Katz, A. N. (1992). Remote memory: recalling autobiographical and public events from across the lifespan. *Can. J. Psychol.* 46, 92–116. doi: 10.1037/h0084311

Hupbach, A., Gomez, R., Hardt, O., and Nadel, L. (2007). Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learn. Mem.* 14, 47–53. doi: 10.1101/lm.365707

James, L. E., Burke, D. M., Austin, A., and Hulme, E. (1998). Production and perception of "verbosity" in younger and older adults. *Psychol. Aging* 13, 355–367. doi: 10.1037/0882-7974.13.3.355

Jansari, A., and Parkin, A. J. (1996). Things that go bump in your life: explaining the reminiscence bump in autobiographical memory. *Psychol. Aging* 11, 85–91. doi: 10.1037/0882-7974.11.1.85

Janssen, S. M. J., Chessa, A. G., and Murre, J. M. J. (2005). The reminiscence bump in autobiographical memory: effects of age, gender, education, and culture. *Memory* 13, 658–668. doi: 10.1080/09658210444000322

Janssen, S. M. J., Kristo, G., Rouw, R., and Murre, J. M. J. (2015). The relation between verbal and visuospatial memory and autobiographical memory. *Conscious. Cogn.* 31, 12–23. doi: 10.1016/j.concog.2014.10.001

Janssen, S. M. J., Rubin, D. C., and St. Jacques, P. L. (2011). The temporal distribution of autobiographical memory: changes in reliving and vividness over the life span do not explain the reminiscence bump. *Mem. Cogn.* 39, 1–11. doi: 10.3758/s13421-010-0003-x

Johnson, M. K., Foley, M. A., Suengas, A. G., and Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *J. Exp. Psychol. Gen.* 117, 371–376. doi: 10.1037/0096-3445.117.4.371

Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. (1993). Source monitoring. *Psychol. Bull.* 114, 3–28. doi: 10.1037/0033-2909.114.1.3

Johnson, M. K., and Raye, C. L. (1981). Reality monitoring. *Psychol. Rev.* 88, 67–85. doi: 10.1037/0033-295X.88.1.67

Kopelman, M. D., Wilson, B. A., and Baddeley, A. D. (1989). The autobiographical memory interview: a new assessment of autobiographical and personal semantic memory in amnesic patients. *J. Clin. Exp. Neuropsychol.* 11, 724–744. doi: 10.1080/01688638908400928

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., and Moscovitch, M. (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychol. Aging* 17, 677–689. doi: 10.1037/0882-7974.17.4.677

McCormack, P. D. (1979). Autobiographical memory in the aged. *Can. J. Psychol.* 33, 118–124. doi: 10.1037/h0081705

Mello, E. W., and Fisher, R. P. (1996). Enhancing older adult eyewitness memory with the cognitive interview. *Appl. Cogn. Psychol.* 10, 403–417. doi: 10.1002/(SICI)1099-0720(199610)10:5<403::AID-ACP395>3.0.CO;2-X

Nadel, L., Samsonovich, A., Ryan, L., and Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus* 10, 352–368. doi: 10.1002/1098-1063(2000)10:4<352::AID-HIPO2>3.0.CO;2-D

Pillemer, D. B. (1992). "Remembering personal circumstances: a functional analysis," in *Affect and Accuracy in Recall: Studies of "Flashbulb" Memories Emory Symposia in Cognition*, Vol. 4, eds E. Winograd and U. Neisser (New York, NY: Cambridge University Press), 236–264.

Piolino, P., Desgranges, B., Benali, K., and Eustache, F. (2002). Episodic and semantic remote autobiographical memory in ageing. *Memory* 10, 239–257. doi: 10.1080/09658210143000353

Rathbone, C. J., Moulin, C. J. A., and Conway, M. A. (2008). Self-centered memories: the reminiscence bump and the self. *Mem. Cognit.* 36, 1403–1414. doi: 10.3758/MC.36.8.1403

Rosenbaum, R. S., Köhler, S., Schacter, D. L., Moscovitch, M., Westmacott, R., Black, S. E., et al. (2005). The case of K.C.: contributions of a memory-impaired person to memory theory. *Neuropsychologia* 43, 989–1021. doi: 10.1016/j.neuropsychologia.2004.10.007

Rubin, D. C. (1982). On the retention function for autobiographical memory. *J. Verbal Learn. Verbal Behav.* 21, 21–38. doi: 10.1016/S0022-5371(82)90423-6

Rubin, D. C. (2000). The distribution of early childhood memories. *Memory* 8, 265–269. doi: 10.1080/096582100406810

Rubin, D. C., and Berntsen, D. (2009). The frequency of voluntary and involuntary autobiographical memories across the life span. *Mem. Cognit.* 37, 679–688. doi: 10.3758/37.5.679

Rubin, D. C., Boals, A., and Berntsen, D. (2008). Memory in posttraumatic stress disorder: properties of voluntary and involuntary, traumatic and nontraumatic autobiographical memories in people with and without posttraumatic stress disorder symptoms. *J. Exp. Psychol. Gen.* 137, 591–614. doi: 10.1037/a0013165

Rubin, D. C., Rahhal, T. A., and Poon, L. W. (1998). Things learned in early adulthood are remembered best. *Mem. Cognit.* 26, 3–19. doi: 10.3758/BF03211366

Rubin, D. C., Schrauf, R. W., and Greenberg, D. L. (2004). Stability in autobiographical memories. *Memory* 12, 715–721. doi: 10.1080/09658210344000512

Rubin, D. C., and Schulkind, M. D. (1997). Distribution of important and word-cued autobiographical memories in 20-, 35-, and 70-year-old adults. *Psychol. Aging* 12, 524–535. doi: 10.1037/0882-7974.12.3.524

Rubin, D. C., and Wenzel, A. E. (1996). One hundred years of forgetting: a quantitative description of retention. *Psychol. Rev.* 103, 734–760. doi: 10.1037/0033-295X.103.4.734

Rubin, D. C., Wetzler, S., and Nebes, R. (1986). "Autobiographical memory across the adult lifespan," in *Autobiographical Memory*, ed. D. C. Rubin (Cambridge: Cambridge University Press), 202–221.

Schrauf, R. W., and Rubin, D. C. (2000). Internal languages of retrieval: the bilingual encoding of memories for the personal past. *Mem. Cognit.* 28, 616–623. doi: 10.3758/BF03201251

St. Jacques, P. L., and Levine, B. (2007). Ageing and autobiographical memory for emotional and neutral events. *Memory* 15, 129–144. doi: 10.1080/09658210601119762

Sutin, A. R., and Robins, R. W. (2007). Phenomenology of autobiographical memories: the memory experiences questionnaire. *Memory* 15, 390–411. doi: 10.1080/09658210701256654

Tulving, E. (1972). "Episodic and semantic memory," in *Organization of Memory*, eds E. Tulving and W. Donaldson (New York, NY: Academic Press), 381–403.

Tulving, E. (1985). Memory and consciousness. *Can. Psychol.* 26, 1–12. doi: 10.1037/h0080017

Waters, T. E. A. (2014). Relations between the functions of autobiographical memory and psychological wellbeing. *Memory* 22, 265–275. doi: 10.1080/09658211.2013.778293

frontiers
in Psychology

Semantic memory as the root of imagination

Anna Abraham[1]* and Andreja Bubic[2]

[1] *School of Social, Psychological and Communication Sciences, Faculty of Health and Social Sciences, Leeds Beckett University, Leeds, UK,* [2] *Psychology Department, Faculty of Humanities and Social Sciences, University of Split, Split, Croatia*

"Imagination is what makes our sensory experience meaningful, enabling us to interpret and make sense of it, whether from a conventional perspective or from a fresh, original, individual one. It is what makes perception more than the mere physical stimulation of sense organs. It also produces mental imagery, visual and otherwise, which is what makes it possible for us to think outside the confines of our present perceptual reality, to consider memories of the past and possibilities for the future, and to weigh alternatives against one another. Thus, imagination makes possible all our thinking about what is, what has been, and, perhaps most important, what might be."—Nigel J. T. Thomas (2004, as cited in Manu, 2006, p. 47)[1].

Investigations of the information processing mechanisms that underlie imaginative thought typically focus on a single branch of imagination, such as prospection, mental imagery or creativity, and are often generalized as being insightful to understanding the workings of imagination in general. In reality, however, there is very little in the way of theoretical or empirical exchange between the scientific communities that conduct research within the different domains of imagination. As a result, the research impetus in each of the sub-domains may be skewed to the pursuit of hypotheses that are not particularly viable in terms of understanding imagination as a whole. An example of this is pegging the roots of imagination to the processes of episodic memory—a reasonable assumption to make based on studies of episodic prospection. However, the associated findings and theoretical conclusions that follow are not entirely consistent with the literature on the mechanisms underlying creativity (Bubić and Abraham, 2014), which is another core realm of imagination.

In an effort to promote interchange across the frontiers of imagination, in this Opinion Article we put forward the idea that all aspects of imagination emerge from semantic memory with increasingly higher-order levels of imaginative information processing emanating from and interacting with existing systems, eventually expanding beyond these to form new systems (**Figure 1**). We compare the associated neurocognitive findings and assumptions in terms of their fit with current knowledge in other fields of imagination and discuss their implications for reformulating hypotheses regarding imagination as a whole.

The What?

Our conceptual knowledge of the world is the foundation from which all imaginative thought emerges and, as such, constitutes "the what-system" within the information processing hub. Investigations of the manner in which concepts are acquired, represented, stored, and accessed fall within the field of semantic cognition. The brain networks that underlie the what-system include

---

[1] http://archive.today/www.imagery-imagination.com or http://www.co-bw.com/BrainConciousness%20Update%20index.htm

FIGURE 1 | An informal illustration of how imaginative processes emerge from and expand beyond semantic memory operations.

modality-specific sensory and motor systems as well as multi-modal or supramodal regions within the inferior parietal lobe, middle and inferior temporal gyri, fusiform and parahippocampal gyri, inferior frontal gyrus, dorsomedial and ventromedial prefrontal cortex and the posterior cingulate gyrus (Binder et al., 2009; Binder and Desai, 2011; Kiefer and Pulvermüller, 2012). Such insights have emerged from neuroscientific investigations into the brain basis of semantic memory, semantic aspects of language processing, and the organization of conceptual knowledge in the brain.

The What–Where?

Determining the location of any object or person relative to one-self, some other person or object is only possible by accessing representations of spatial information such as direction, orientation, distance and position of that object or person. Such information is coded by means of reference frames relative to the observer (egocentric) and independent of the observer (allocentric) (Burgess, 2006). Tasks of spatial memory and navigation have shown that medial temporal lobe structures such as the hippocampal formation, parahippocampal gyrus, entorhinal and perirhinal cortices as well as medial parietal regions, such as the retrosplenial and posterior cingulate cortices (Burgess, 2008; Chrastil, 2013; Ekstrom et al., 2014), are critically involved in

spatial information processing. Others tasks of spatial cognition, such as perspective taking, have indicated the involvement of additional regions within the posterior parietal cortex, particularly the inferior parietal and temporo-parietal areas (Byrne and Becker, 2007; Dhindsa et al., 2014).

The What–Where–When?

An event is defined as a specific happening (what) that occurs at a certain place (where) and at a given time (when). During retrospection we access events from our personal past (episodic memory, autobiographical memory), whereas during prospection we contemplate events that could unfold in our personal future (episodic future thinking). Both fall within the umbrella concept of mental time travel (Tulving, 1985). Neuroscientific evidence has consistently revealed that the brain network that is engaged when we imagine personal events in the near or distant future overlaps considerably with the network that is activated when we ponder our episodic or autobiographical past (Schacter et al., 2007, 2012; Mullally and Maguire, 2013). Regions that comprise this brain network include the ventral and dorsal medial prefrontal cortex, retrosplenial and posterior cingulate regions within the medial parietal cortex, anterior lateral temporal cortex, inferior parietal cortex, and medial temporal lobe structures such as the hippocampus. Notably, the

regions of the mental time travel brain network also closely correspond to those of the brain's default mode network (DMN). The DMN is active under conditions of rest and low task load, and is held to reflect processing demands associated with mind-wandering, internal mentation and stimulus-independent thought (Andrews-Hanna et al., 2014). DMN brain areas are also involved in other facets of higher order cognition, like mental state reasoning or theory of mind, moral cognition, and self-referential thought (Buckner et al., 2008; Spreng et al., 2009), all of which involve reasoning about one's self and/or others.

The What-If?

Our capacity to imagine possibilities is virtually unconstrained. Investigations on the information processing circuits involved in prospection address the question of "what if?" or "what might be?" within a specific temporal context of our personal lives (which covers the aforementioned episodic prospection of the what–when–where system). However, our cognitive capacity to explore hypothetical possibility spaces is neither limited only to our personal lives nor to any temporal factor (past/present/future). Other operations that fall under the category of what-if or hypothetical reasoning based cognitive processes include semantic prospection, semantic or episodic counterfactual reasoning and creativity. In addition to the partial conceptual overlap between the what-if system and the previously discussed what-where-when system, the two also share common underlying neural mechanisms. Although only a few neuroscientific studies have investigated semantic prospection or the propensity to think about the non-personal future, the limited evidence indicates that semantic prospection is reliant on similar parts of the brain's episodic mental time travel network, particularly with reference to the engagement of anterior and dorsal medial prefrontal regions, inferior parietal cortices, the hippocampus and related medial temporal lobe structures (Abraham et al., 2008; Race et al., 2013).

In contrast to semantic prospection, which is relatively unrestricted with regard to the types of imaginable alternatives, counterfactual thinking primarily involves exploring possibilities that are contrary to what has already come to pass. Research on brain correlates of counterfactual comparisons and emotions that often accompany such cognition, such as regret, indicates a key role for the orbitofrontal and ventromedial prefrontal cortices (Camille et al., 2004; Levens et al., 2014). Furthermore, studies that have assessed episodic past, episodic future and episodic counterfactual thinking have reported a common brain network, involving the hippocampal formation, temporal lobe structures, lateral parietal regions as well as medial and lateral prefrontal areas. Within the episodic cognition domain, counterfactual thinking recruits some of these areas more strongly than past and future thinking, and also additionally engages the bilateral inferior parietal lobe and posterior medial frontal cortex (Van Hoeck et al., 2013).

Semantic prospection and counterfactual reasoning are concerned with hypothetical reasoning linked to the future and the past, respectfully. However, one can also engage in hypothetical reasoning within temporally unspecific contexts such as those involving moral and mental state reasoning, which, as pointed out earlier, strongly overlap in terms of their implicated brain network with the what–when–where system (Buckner et al., 2008). While the contexts tapped in such hypothetical reasoning operations are decidedly social in nature, a non-socially based avenue within which we necessarily exercise our capacity to think hypothetically is that of creativity.

Our capacity to be creative is examined by assessing the extent to which we are able to generate original and relevant responses to a particular end (Stein, 1953; Runco and Jaeger, 2012). The underlying brain mechanisms of creative cognition are very complex (Abraham, 2014). Brain regions such as the dorsal and ventral medial prefrontal cortex, retrosplenial and posterior cingulate cortices as well as medial temporal lobe structures are strongly engaged during divergent thinking, or the generation of multiple responses in an open-ended situation (Abraham et al., 2012). This indicates that there is a considerable overlap in the neural correlates of divergent thinking and that of the what–when–where network. While divergent thinking certainly involves hypothetical reasoning and exploration of an abstract possibility space, it does not necessarily translate to creative thought. Having constraints on divergent thinking pushes the information processing system to be necessarily creative (both original and relevant) and this leads to the additional activation of the semantic cognition and cognitive control networks with the major contributions being provided by brain regions such as the inferior frontal gyrus, temporal pole, frontopolar cortex, and basal ganglia. So the neural correlates of creative cognition system overlap only partially with those associated with other aspects of the imagination system with common activations seen in the dorsomedial prefrontal cortex and inferior parietal lobe (the what–when–where system) as well as the inferior frontal gyrus (the what-system).

Integrating the Disparate Systems of Imagination

In this Opinion Article, we explored the view that processes of imagination—the "where" of spatial cognition, the "what-when-where" of episodic retrospection and prospection, and the "what-if" of semantic prospection, counterfactual reasoning and creative thinking—emerge from a foundation provided by the "what" of semantic memory operations. The evidence thus far clearly indicates that the many processes of imagination, which have mostly been systematically investigated in isolation from one another, are neurally implemented in substantially overlapping brain networks and are also similar with respect to their underlying cognitive algorithms and mechanisms. This resonates with other proposals that have highlighted that semantic and episodic cognitive operations and their related brain systems are dynamically interlinked (Squire and Zola, 1998; Greenberg and Verfaellie, 2010), as well as with recent calls for de-emphasizing the episodic or autonoetic aspects of future oriented cognition and advocating the central role played by semantic memory in the same (Stocker, 2012; Irish and Piguet, 2013).

This does not mean that all imaginative processes are to be considered "atemporal" per definition. Many forms of mental time travel as well as counterfactual thinking patently involve the consideration of temporal factors as a core facet of the imaginative process. In taking this a step further, it may even be argued that such processes are necessarily linked to the brain's predictive systems due to the fact that they involve the generation of estimates concerning events that reliably unfold over a certain period of time, albeit with differing levels of certainty (Bubic et al., 2010). This position has rarely been considered in the literature on imagination-relevant operations but it would fit with a number of suggestions that posit prediction as the fundamental mechanism that modulates our general neural and cognitive processing (Friston and Stephan, 2007; Pezzulo, 2008).

So, although the issue of temporality is undoubtedly relevant, the more fundamental basis that underlies all of the aforementioned processes is the reliance on our experiences with the world, its objects and events. We therefore suggest that if the aim is to develop a comprehensive information processing model of imagination, the foundational elements should be discussed in terms of semantic memory operations. As semantic memory involves the abstraction of content from experiences that are specific to sensory, motor, or affective modalities, conceptualizing the processes of imagination as stemming from semantic operations allows for a more seamless integration of its theoretical models with that of the wider research realm of perception, action and cognition where concepts such as embodied cognition and predictive processing are revolutionizing our understanding of psychology.

We hope these ideas will stimulate future research and the development of novel paradigms as well as critical scientific exchange between the research communities involved in understanding different aspects of imagination. Some questions can be already anticipated such as the "chicken-and-egg" problem within which it appears impossible to clearly substantiate what came first, or concerns about how to reach a consensus about what can be considered an underlying foundational element. Through the process of this discussion though, we hope that building blocks and essential frameworks will be uncovered that will guide us through the incredibly rich world of human imagination.

References

Abraham, A. (2014). Creative thinking as orchestrated by semantic processing vs. cognitive control brain networks. *Front. Hum. Neurosci.* 8:95. doi: 10.3389/fnhum.2014.00095

Abraham, A., Pieritz, K., Thybusch, K., Rutter, B., Kröger, S., Schweckendiek, J., et al. (2012). Creativity and the brain: uncovering the neural signature of conceptual expansion. *Neuropsychologia* 50, 1906–1917. doi: 10.1016/j.neuropsychologia.2012.04.015

Abraham, A., Schubotz, R. I., and von Cramon, D. Y. (2008). Thinking about the future versus the past in personal and non-personal contexts. *Brain Res.* 1233, 106–119. doi: 10.1016/j.brainres.2008.07.084

Andrews-Hanna, J. R., Smallwood, J., and Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann. N.Y. Acad. Sci.* 1316, 29–52. doi: 10.1111/nyas.12360

Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536. doi: 10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055

Bubić, A., and Abraham, A. (2014). Neurocognitive bases of future oriented cognition. *Rev. Psychol.* 21, 3–15. doi: 10.3389/fnhum.2010.00025

Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.* 4:25. doi: 10.3389/fnhum.2010.00025

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N.Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends Cogn. Sci.* 10, 551–557. doi: 10.1016/j.tics.2006.10.005

Burgess, N. (2008). Spatial cognition and the brain. *Ann. N.Y. Acad. Sci.* 1124, 77–97. doi: 10.1196/annals.1440.002

Byrne, P., and Becker, S. (2007). Remembering the past and imagining the future. *Psychol. Rev.* 114, 340–375. doi: 10.1037/0033-295X.114.2.340

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., and Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304, 1167–1170. doi: 10.1126/science.1094550

Chrastil, E. R. (2013). Neural evidence supports a novel framework for spatial navigation. *Psychon. Bull. Rev.* 20, 208–227. doi: 10.3758/s13423-012-0351-6

Dhindsa, K., Drobinin, V., King, J., Hall, G. B., Burgess, N., and Becker, S. (2014). Examining the role of the temporo-parietal network in memory, imagery, and viewpoint transformations. *Front. Hum. Neurosci.* 8:709. doi: 10.3389/fnhum.2014.00709

Ekstrom, A. D., Arnold, A. E. G. F., and Iaria, G. (2014). A critical review of the allocentric spatial representation and its neural underpinnings: toward a network-based perspective. *Front. Hum. Neurosci.* 8:803. doi: 10.3389/fnhum.2014.00803

Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

Greenberg, D. L., and Verfaellie, M. (2010). Interdependence of episodic and semantic memory: evidence from neuropsychology. *J. Int. Neuropsychol. Soc.* 16, 748–753. doi: 10.1017/S1355617710000676

Irish, M., and Piguet, O. (2013). The pivotal role of semantic memory in remembering the past and imagining the future. *Front. Behav. Neurosci.* 7:27. doi: 10.3389/fnbeh.2013.00027

Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* 48, 805–825. doi: 10.1016/j.cortex.2011.04.006

Levens, S. M., Larsen, J. T., Bruss, J., Tranel, D., Bechara, A., and Mellers, B. A. (2014). What might have been? The role of the ventromedial prefrontal cortex and lateral orbitofrontal cortex in counterfactual emotions and choice. *Neuropsychologia* 54, 77–86. doi: 10.1016/j.neuropsychologia.2013.10.026

Manu, A. (2006). *The Imagination Challenge: Strategic Foresight and Innovation in the Global Economy.* Berkeley, CA: New Riders.

Mullally, S. L., and Maguire, E. A. (2013). Memory, imagination, and predicting the future: a common brain mechanism? *Neuroscientist* 20, 220–234. doi: 10.1177/1073858413495091

Pezzulo, G. (2008). Coordinating with the future: the anticipatory nature of representation. *Minds Machines* 18, 179–225. doi: 10.1007/s11023-008-9095-5

Race, E., Keane, M. M., and Verfaellie, M. (2013). Losing sight of the future: impaired semantic prospection following medial temporal lobe lesions. *Hippocampus* 23, 268–277. doi: 10.1002/hipo.22084

Runco, M. A., and Jaeger, G. J. (2012). The standard definition of creativity. *Creat. Res. J.* 24, 92–96. doi: 10.1080/10400419.2012.650092

Schacter, D. L., Addis, D. R., and Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* 8, 657–661. doi: 10.1038/nrn2213

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., and Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron* 76, 677–694. doi: 10.1016/j.neuron.2012.11.001

Spreng, R. N., Mar, R. A., and Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J. Cogn. Neurosci.* 21, 489–510. doi: 10.1162/jocn.2008.21029

Squire, L. R., and Zola, S. M. (1998). Episodic memory, semantic memory, and amnesia. *Hippocampus* 8, 205–211.

Stein, M. I. (1953). Creativity and Culture. *J. Psychol.* 36, 311–322. doi: 10.1080/00223980.1953.9712897

Stocker, K. (2012). The time machine in our mind. *Cogn. Sci.* 36, 385–420. doi: 10.1111/j.1551-6709.2011.01225.x

Tulving, E. (1985). *Oxford Psychology Series (New Ed Edn)*. Oxford: OUP Oxford.

Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., Vandekerckhove, M., and Van Overwalle, F. (2013). Counterfactual thinking: an fMRI study on changing the past for a better future. *Soc. Cogn. Affect. Neurosci.* 8, 556–564. doi: 10.1093/scan/nss031

ADVANTAGES OF PUBLISHING IN FRONTIERS

FAST PUBLICATION

Average 90 days
from submission
to publication

**COLLABORATIVE
PEER-REVIEW**

Designed to be rigorous –
yet also collaborative, fair and
constructive

RESEARCH NETWORK

Our network
increases readership
for your article

OPEN ACCESS

Articles are free to read,
for greatest visibility

TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles

GLOBAL SPREAD

Six million monthly
page views worldwide

COPYRIGHT TO AUTHORS

No limit to
article distribution
and re-use

IMPACT METRICS

Advanced metrics
track your
article's impact

SUPPORT

By our Swiss-based
editorial team