NEUROSCIENCE of Consciousness

Neuroscience of Consciousness, 2017, 1–13

doi: 10.1093/nc/nix006
Research article

Meta-analytic comparison of trial- versus questionnaire-based vividness reportability across behavioral, cognitive and neural measurements of imagery

Matthew S. Runge[1], Mike W.-L. Cheung[2] and Amedeo D'Angiulli*,[1]

[1]Department of Neuroscience, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6; [2]Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Block AS4 Level 2, 9 Arts Link, Singapore 117570

*Correspondence address. Department of Neuroscience & Institute of Interdisciplinary Studies, Carleton University, 1125 Colonel By Drive, Dunton Tower, Ottawa, Ontario K1S 5B6, Canada. Tel: (613) 520-2600, ext. 2954 (office), (613) 520-2600, ext. 3097 (lab); Fax: (613) 520-2301; E-mail: amedeo.dangiulli@carleton.ca

Abstract

Vividness is an aspect of consciousness related to mental imagery and prospective episodic memory. Despite being harshly criticized in the past for failing to demonstrate robust correlations with behavioral measures, currently this construct is attracting a resurgent interest in cognitive neuroscience. Therefore, an updated examination of the validity of this construct is timely. A corpus of peer-reviewed literature was analyzed through meta-analysis, which compared the two main formats used to measure vividness [trial-by-trial vividness ratings (VR) and the Vividness of Visual Imagery Questionnaire (VVIQ)]. These two formats were compared in relation to all available behavioral/cognitive (BC) and neuroscience (NS) measures in Phase 1 (3542 statistical observations representing 393 journal articles); and then in relation to all available BC, EEG and fMRI literature in Phase 2 (3624 observations representing 402 articles). Both Phases observed significantly larger effect size estimates (ESEs) for VR than VVIQ, and larger ESEs for NS than BC measures. ESEs for EEG and fMRI were not significantly different in Phase 2, but were greater than BC ESEs. These data suggest VR are a more reliable self-report measure than VVIQ, and may reflect a more direct route of reportability than the latter. Furthermore, both VR and VVIQ are more strongly associated with the neural, than the cognitive and behavioural correlates of imagery. If one establishes neuroscience measures as the criterion variable, then self-reports of vividness show higher construct validity than behavioural/cognitive measures of imagery. We discuss how the present findings contribute to current issues on measurement of reportability; and how this study advances our understanding of vividness as a phenomenological characteristic of imagery, and other forms of conscious experience which do not necessarily involve imagery.

Key words: vividness; vividness of visual imagery questionnaire; imagery; neuroimaging; validity

Introduction

Visual mental imagery refers to the subjective experience of a percept-like pattern in the absence of a relevant physical stimulus on the retina (Hebb 1968; D'Angiulli 2008). Historically, Galton (1880) was the first to observe that the "detail and clarity with which individuals experience mental imagery" involves an individual difference gradient across a population, which he

Revised (in revised form): 24 February 2017. **Accepted:** 20 March 2017

operationally defined as "vividness." While much progress has been achieved in clarifying the neural and information processing nature of mental imagery, dimensions such as vividness, which concern the phenomenal experience of imagery, have potentially been neglected by contemporary cognitive psychology. Indeed, after Marks (1999), in the last 15 years very few examples can be found in the literature attempting to readdress this topic and its ambiguousness [see D'Angiulli *et al.* (2013)] or situate it within the broader arena of consciousness research [see Andrade (2001)]. Correspondingly, it seems very little progress has been made in terms of a complete account of mental imagery which offers a constructive integration of its component dimensions – cognitive, affective, neural, and phenomenological [the term "constructive" is borrowed from Flanagan and Dryden (1998)].

Lack of progress in this direction is surprising, given that no author ever denies the experiential aspects of imagery (or, similarly, that vividness of imagery is relevant to the scientific study of inner experience). On the contrary, experiential aspects of mental imagery continue to be used circularly, and ubiquitously for defining the psychological status and relevance of mental imagery [for both points, see Pylyshyn (2003)]. For example, imagery paradigms that require subjects to provide an observable objective behavior (e.g. a button press) in response to generating, holding or transforming images rely on subjects being aware that they are experiencing an image. Thus, objective responses during imagery tasks are incontrovertibly entangled with the report of phenomenological awareness [see Overgaard (2006)]. Because conscious phenomenal awareness of the imagery experience is constitutive of what is reported (the content of conscious experience), the button press or other behavioral measures are not independent of the reported content (Georgalis 2006). In other words, there is no independent way of telling whether the button press or other behavioral measures involve visual mental imagery, unless the latter experience is explicitly reported by the subject. This type of circularity is generally acknowledged in the broader context within the study of visual consciousness (Naccache 2008; Dehaene 2014).

In the "imagery debate," two contending approaches provide the best accounts for several aspects of imagery, and still remain the most influential cognitive psychology approaches to its study. According to Kosslyn (1994)'s "pictorial theory," visual mental images are constructed from depictive representations, which come into play at higher stages of visual information processing during actual perception. According to Pylyshyn (2003)'s "tacit knowledge," experimental data on visual mental images simply reflect different instances of what people implicitly know or believe about how they perceive. Although Kosslyn and Pylyshyn do not explicitly acknowledge, or refer to the vividness construct overtly in their theories, and both authors might even summarily dismiss the relevance of the vividness construct, they allude to it at several critical junctions in their explanatory arguments, as evidenced by the following examples.

In discussing the mental scanning paradigm and the size of mental images, to illustrate what tacit knowledge involves, Pylyshyn (2003, p. 163) repeatedly refers to "details" represented in images and "blurry" versus "clear" images. These adjectives epitomize the two terms used to describe vividness, according to the classical definition of vividness in psychology (Galton 1880, 1883; Betts 1909; note the historical reference has been introduced by the authors of the current paper, Pylyshyn did not mention it). Consequently, at least some of his key conclusions about self-reports assume the validity of the vividness measure is sufficient.

On the other hand, in a paper that set the methodology benchmark for a series of fMRI studies showing activation of V1 by visual imagery, Kosslyn *et al.* (1999) use the word vividness when describing their mental imagery model based on Hebb's cell assembly hypothesis; especially, in framing some aspects of their data. For example, when discussing the controversy concerning the reinterpretation of mental images (Reisberg 1996), these authors argue that past imagery reinterpretation tasks may have required high-resolution images. Cues were effective in facilitating reinterpretation because their "... images were not very vivid" (Kosslyn *et al.* 1999, p. 286). Additionally, these authors defend another conclusion carefully stating that "... additional effort is required in imagery to represent visual patterns with high resolution." However, resolution and vividness were equated earlier in the article: "translated in our present understanding, more vivid images would occur when stored information activates lower areas, which have higher spatial resolution than do higher areas" (Kosslyn *et al.* 1999, p. 277). Correspondingly, at least some of the supporting explanations used by Kosslyn *et al.* go circularly back to vividness.

Even if these authors do not acknowledge vividness as a classical and contemporary psychological construct, it is nonepiphenomenal in the context of their explanations, and cannot be eliminated, nor refuted at present. In addition, a number of imagery researchers have shown that properties available to self-report, such as the ones typically included in definitions of vividness can reflect the resolution of the visual buffer [see Dean and Morris (2003); D'Angiulli (2002)]. Furthermore, at the crossroads with imagery – episodic memory and cognitive neuroscience – some investigators use self-report variables such as Paivio's imageability (Gonsalves *et al.* 2004), or other contextual variables in their experiments (Wheeler *et al.* 2000) as an indirect way to observe, interpret, and discuss the underlying vividness of the cognitive event. Thus, it seems that contentions toward the validity of the vividness construct do not take in consideration subtle differences in how vividness is actually measured, or what vividness actually means. Reexamination of the construct of vividness is timely also because multiple versions of essentially a similar latent construct are increasingly being used in different areas of neuroscience and psychology related to, but other than mental imagery, such as prospective and episodic memory (St Laurent *et al.* 2015), and aging (Johnson *et al.* 2015). Vividness is increasingly becoming a central topic for phenomenological, neurobiological, and genetic links between visual memory and emotion (Markovic *et al.* 2014). Thus, the concept has the potential to advance many fields of research, but presumably for structural and historical constraints, transdisciplinary integration of advances in knowledge seems to progress inductively, from more specific, particular topics (case in point, vividness in imagery) to the general (consciousness).

With such transdisciplinary background in mind, the purpose of the present article is to update the status of vividness. We attempt to achieve the goal by working toward a new way to substantiate the construct with empirically demonstrated validity, which includes current findings from neuroscience, among other different disciplines. Our approach to the study of the validity of vividness, however, is based on "reference" rather than "meaning" (Borsboom 2005). From this perspective, the crucial issue is not what "vividness really means," but rather, to what extent our subjective measures of imagery vividness work; namely, whether they measure the intended core latent attribute of inner conscious experience that comes with having visual mental images. Therefore, we already make the prior assumption that there does exist, in reality, an attribute that one designates when using the term "vividness." Following this approach, the issue of validity can be reduced to whether a measurement instrument is sensitive to variations of the assumed attribute (Borsboom *et al.* 2004). Namely, the question becomes: what do subjective vividness instruments really measure?

To find out what is measured, one has to find out how the instrument works. Accordingly, the development of any construct requires the refinement of its operational definition, namely, defining the range of variation of the underlying assumed attribute, such that it can be calibrated more precisely against related psychological and neurophysiological variables (Markus and Borsboom 2013). For example, to illustrate one of the most relevant ambiguity that has been identified over the years, although vividness is operationally defined by the key words "detail" and "clarity," Kensinger *et al.* (2011) observed a neurophysiological distinction between the vividness of a memory and the detail a memory possesses, as they relate to brain activation patterns. Such ubiquity could be interpreted as contradictory, but from the empirical point of view just confirms earlier phenomenological findings, that across several different instruction conditions, people seem to heed to, and report on, both an intensive aspect of the sensory strength of their images (or "clarity," generally described as brightness and color) and the level of precision or "detail" discerned within them (Kosslyn and Alper 1977; McKelvie 1994; Marks 1999).

As two very recent landmark studies show, the ubiquity that seems associated with the construct of vividness almost invariably reflects empirical variations in individual differences. On the one hand, it has been shown the overlap in neural activation (especially in the early primary visual cortex, V1) between imagery and perception is directly, positively correlated with trial by trial vividness ratings (VR) and imagery ability (Dijkstra *et al.* 2017). On the other hand, individual differences in the imagery experience are found to vary according to the two inversely correlated attributes associated with the size of V1. Individuals with above-average V1 size experience higher detailedness but lower sensory strength, whereas individuals with below-average V1 size show the opposite pattern (Bergman *et al.* 2016). But most important, performance and verbal reports by the majority of people, with average V1 size, reflect a relative mixture of those two visual features of images.

Previous phenomenological and behavioral (Reaction Times modeling) evidence confirms the trend in the majority of the observers, and based on the weight of this evidence combined with neurophysiological data (Baddeley and Andrade 2000; D'Angiulli and Reeves 2007) some authors defend that it seems reasonable to assume as a definition of vividness the most exhaustive subjective conscious experience of imagery, in terms of both amount of detail (resolution) and sensory strength of a mental image, relative to the experience of actual seeing. The latter operational working definition is the frame of reference for the tests of validity proposed in this article.

A line of thinking about validity similar to the one just discussed has led to some of the strongest arguments demarking the necessity and objectifiability of self-reports in cognitive neuroscience, and consciousness research more generally. Particularly relevant is Jack and Roepstorff's (2003) analysis of the "triangulation" in consciousness research; namely, the convergent use of introspective, behavioral and neurophysiological evidence. As they argued, empirical validity of introspective reports cannot be only established by examining the relationship between reported vividness and other objective behavioral measures (e.g. memory accuracy). Rather, empirical validity should be based on the relationship between neural correlates of perceived vividness and reported vividness at the first-person level. Without considering perceived vividness (what the person is actually experiencing), the fact that reported vividness may not correlate with behavior does not refute the validity of the vividness construct. Crucially, reported vividness may still be strongly correlated with the neural correlates of perceived vividness experience. Because vividness was defined ∼130 years ago, and its main synthetic assessment [the meta-analysis by McKelvie (1995)] did not include current neuroscience findings, the construct requires an updated reconsideration as to what it actually means in the fields of consciousness and imagery at present. This effort may accommodate novel, contemporary conceptualizations concerning the nature of memory (Moscovitch *et al.* 2005), or even new knowledge regarding the neural basis of consciousness, such as the default mode network (Andrews-Hanna *et al.* 2010).

In addition, as a psychological construct, vividness requires convergent and discriminative validation through empirical, objective testing. For instance, vividness has been positively correlated with performance on perceptual and memory tasks (Baddeley and Andrade 2000; Rodway *et al.* 2006; Lee *et al.* 2012), arousal level (Barrowcliff *et al.* 2004; Bywaters *et al.* 2004), and sleep stages (Conduit *et al.* 2004), even though it is debated whether or not it is correlated with certain dynamic spatial tasks like mental rotation (Dean and Morris 2003; Logie *et al.* 2011). Moreover, a growing body of research suggests VR correlate with neural modulations in specific brain regions (Gilboa *et al.* 2004; Cui *et al.* 2007; Logie *et al.* 2011; Cichy *et al.* 2012; Todd *et al.* 2013), which covary with the phenomenological ratings offered by participants.

Perhaps the most commonly used global assessment of imagery ability is the Vividness of Visual Imagery Questionnaire (VVIQ; Marks 1973; Pearson 1995; Lee *et al.* 2012) and its successor the VVIQ2 (Marks 1995; Shen *et al.* 2015). The VVIQ provides a global assessment of vividness, and is typically used to categorize participants according to the individual differences in visual imagery ability. It consists of 16 items, which are to be rated on a five-point scale from (1) "perfectly clear and as vivid a normal vision," to (5) "no image at all, you only 'know' that you are thinking of the object." The versatility of the VVIQ is such that it can be administered before, during, or after experimental manipulations, and demonstrates sufficient retest and internal reliability (Campos *et al.* 2002).

Alternatively, vividness may be rated on a "trial-by-trial" VR basis through a single response, which corresponds to the subjective experience at a particular moment in time, structurally constrained by what types of images participants are required to form. In addition to methodological advantages, Hertzog and Dunlosky (2006) argue that trial-by-trial VR are perhaps the most effective means by which imagery vividness can be studied (e.g. D'Angiulli and Reeves 2003–2004). Similarly, participants are prompted with a scale such as that used in the VVIQ, wherein vividness is rated from (1) "no image," to (5) "perfectly vivid" (albeit the scale can range from 1 to 3, 1 to 7, 1 to 100, etc.). A major contention concerning the use of VVIQ is that global assessments offered through surveys of group differences in mental imagery ability may describe too coarsely, or even miss specific cognitive and neural processes affiliated with the phenomenological experience of vividness occurring within each trial (D'Angiulli *et al.* 2013).

The reliability of vividness, as a construct presumed to reflect identifiable and separable processes, is largely undermined through the use of a single test score, such as the one offered by the VVIQ, or global assessments of imagery ability more generally (Pearson 1995). In contrast, experimental procedures employing VR may procure more reliable patterns of results and interpretations than those employing the VVIQ (D'Angiulli 2002; D'Angiulli and Reeves 2002, 2007; Alter and Balcetis 2010; Rabin *et al.* 2010; Pearson *et al.* 2011). If VR are observed in structured experimental settings, and participants are given clear instructions as to the nature of the task, VR may resolve critically informative and specific aspects concerning the subjective experience of the imagery processes (Baddeley and Andrade 2000). In this way, VR are a type of retrospective verbal report (Ericsson and Simon 1993), which can be put in correspondence with decomposable task-related behavior and neural processes. Accordingly, VR is compatible with the principle of "reportability," i.e., the more general condition with which the levels and contents of phenomenal conscious awareness in visual working memory can be properly defined and investigated (Weiskrantz *et al.* 1995; Naccache 2008).

Experimental results which are based on behavioral, cognitive, or neuroscience types of measures are assumed to produce effects which vary in their relative sizes; however, there has been no systematic quantitative comparison that evaluates the claim, nor offers an evidence-based indication as to whether vividness provides weak or robust data [for a complete account of this debate, see McKelvie (1995)]. Adopting a meta-analysis approach offers the opportunity to formulate a set of clear and straightforward hypotheses and predictions that could be put forward to empirical testing. Because VRs resolve critical aspects of, and measure a greater subset of the variability in the mental imagery experience, they may systematically demonstrate greater content validity than the VVIQ within behavioral, cognitive, and neuroscience experimental paradigms (Hale and Simpson 1971; Giusberti *et al.* 1992; Baddeley and Andrade 2000; Smeets *et al.* 2012). Accordingly, one would expect larger effect sizes for VR [see Richardson (1994)]. If VR resolve phenomenological aspects of mental imagery to a greater degree than the VVIQ (i.e. they are more exhaustive), then the reported effect sizes pertaining to VR should be greater in magnitude than the effect sizes associated with the VVIQ, when averaged over a large and representative sample of relevant literature. To test this first key hypothesis, in the present meta-analysis, VR were compared against all available behavioral, cognitive, and neuroscience imagery measures throughout a robust and representative sample of literature with the VVIQ providing the criterion variable against which self-reports are traditionally compared and validated (McKelvie 1995).

The viability of an evidence-based approach to mental imagery is one which is theoretically driven by and supported through neurocomputational modeling (Cichy *et al.* 2012). In addition, neuroimaging experiments typically exhibit a great amount of control and precision (Kosslyn 1994). Thus, the second hypothesis tested through the present meta-analysis was that neuroscience measures may resolve critical aspects of the mental imagery experience more reliably than behavioral and or cognitive ones, and as result, it would be possible to conclude that vividness may be more thoroughly validated as an independent and empirical construct through neuroscience approaches. If neuroscientific measures are capable of resolving phenomenological vividness more closely than behavioral and or cognitive ones, then the prediction follows that the reported magnitude of effect sizes which are neuroscience should be greater than those which are behavioral and or cognitive. Specifically, VRs should also be greater than the VVIQ for effect sizes which are associated with neuroscience measurements. The latter outcome would entail a broader implication in terms of how conscious verbal reports of imagery could be conceptualized. That is, imagery reportability may be associated with two routes, a "direct" route, represented by VRs, and an indirect one, represented by VVIQ. Following Ericsson and Simon's (1980) general typologies of verbal protocols, it would be reasonable to interpret the predicted pattern as an indication that VR and VVIQ may reflect two different processing routes supporting retrospective verbal reports as postulated in the schemas shown in Fig. 1. Like other types of retrospective reports, VR are obtained by asking subjects about a neurocognitive process which recently occurred. It is reasonable that in structured conditions, in which VR are collected, the image contents may be accessed through a relatively more direct route, with minimal verbal recoding and delay that reflect predominantly or exclusively task-relevant information and processes (Fig. 1, right panel). Conversely, in the case of VVIQ, VR may require additional scanning, filtering, inference or rule-based and metacognitive processes; as a result, verbalization may mediate or even modify the imagery experience revealing information or knowledge that goes beyond or is not relevant to processes underlying the task (Fig. 1, left panel).

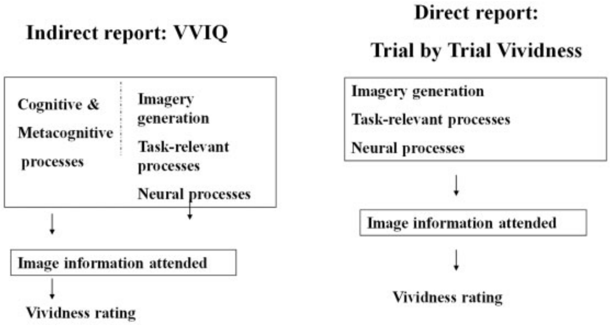

Figure 1. Schematic models representing types of VR (left panel: VVIQ; right panel: trial-by-trial VR) as particular cases of two retrospective verbal report processing routes, adapted from Ericsson and Simon (1980). Like other types of retrospective reports, VR are obtained by asking subjects about a process which occurred recently. This figure shows the cases similar to introspection, in which VR requires scanning, filtering, inference, or rule-based processes, thus verbalization mediates or even modifies the imagery task.

A third hypothesis may also be derived as a possible contingency from the first two, if the presence of a significant interaction is observed between Vividness and Measure factors, it may represent a differential propensity for VR/VVIQ within the behavioral/cognitive (BC) and/or NS dimensions. Because the effect sizes for VR are hypothesized to exceed those of the VVIQ for both BC and NS measurements, failing to reject the null hypothesis for the interaction effect would imply the strongest form of convergent validation, and support for the first two hypotheses.

Method: Phase 1

Sampling

A corpus of peer-reviewed journal articles representing a robust subset of the relevant literature available through the databases Web of Science, Scopus, Embase, and PsycINFO were retrieved through our University's library on October 25th, 2012. *A priori* criteria restricted the search results to those of the English language, to those published after 1950, and those using human subjects. Given the current major issues concerning the inclusion of "grey literature" – such as poorer research quality, search un-standardization, data duplication and/or incompleteness, minimal reliability improvement at cost of major energy/resource search expenditure [see extensive review in Hopewell *et al.* 2005] – theses, dissertations, and other unpublished works were deliberately avoided (a more detailed rationale for this decision contextualized to the present study is given in the "Discussion" section). Data for Phase 1 are available upon request.

Search parameters included the general terms "vivid* and image*", as well as additional syntax employed to minimize the discovery of irrelevant papers (see Supplementary Material). A total of 1290 journal articles were discovered (Web of Knowledge, 36; Scopus, 34; Embase, 188; PsycINFO, 1032), and exported to RefWorks, wherein 320 duplicates were observed and deleted from the corpus. The remaining 970 journal articles were systematically analyzed in relation to their relevance to the research question. Those which did not include at least one relevant statistical outcome relating either vividness ratings (VR) and/or VVIQ to another variable were not considered further, of which there were 577. From the remaining 393 papers that contained at least one relevant statistical outcome, each and every statistical outcome relating either VR and/or VVIQ to another variable was recorded into a database, as it appeared in the original journal article.

Data trimming and data analysis

Upon inclusion into the database, every statistical outcome was categorized as either VR or VVIQ, and further categorized as either BC, or neuroscience (NS). A total of 3697 statistical outcomes were observed from the 393 papers included, which were organized into four focal categories (VR_{BC}, $n = 1826$; VVIQ_{BC}, $n = 1680$; VR_{NS}, $n = 82$; VVIQ_{NS}, $n = 62$), and one peripheral category consisting of observations which directly correlated VR with VVIQ ($n = 47$).

Each statistical outcome was transformed from its original statistical metric into the form of the correlation coefficient (r), from which absolute Fisher's Z_r transformed score were computed. Statistical outcomes for which r could not be calculated were excluded from the analysis ($n = 108$), such as beta-values ($n = 73$), t, F, r, U, and q values which did not included sufficient information ($n = 35$). Because scaling phenomenological ratings from 1 (low) to 5 (high), has the exact opposite meaning as scaling phenomenological ratings from 5 (low) to 1 (high), rearranging statistical outcomes to accommodate the scaling *post priori* was methodologically difficult. Subsequently, the present research abandoned directionality and transformed Z_r vector quantities to their absolute, scalar Z_r form. Observations directly measuring VR and VVIQ together were considered separately, such that an average correlation could be isolated.

Data were modeled with a series of three-level meta-analysis (Van den Noortgate *et al.* 2013; Cheung 2014). Traditional meta-analytic approaches assume independence in the effect sizes (Hox *et al.* 2010). As there are 3697 effect sizes nested within 393 studies in the present meta-analysis, statistical inferences of traditional meta-analysis are incorrect. Three-level meta-analysis enables researchers to implement an additional cluster effect (dependence within the same study). Although ignoring dependence is not recommended, aggregating multiple statistical outcomes into one summary statistic can likewise be problematic (Hedges and Pigott 2001), as it may afford larger standard errors affiliated with parameter estimates, and generally contributes to a high attrition rate (loss of sample size). As such, the present research necessitated a more precise and accommodating meta-analytic framework to answer the primary research question in Phase 2. Data were analyzed using R software, with the "metaSEM" package (Cheung 2015). The level-2 and level-3 heterogeneity variances represent the within- and between-study heterogeneity variances, respectively. If the level-2 heterogeneity variance is large, it means that the reported effect sizes also vary within the same study.

Correspondingly, data for VR_{BC} ($n = 1760$) represented 238 experiments from 194 journal articles, VVIQ_{BC} ($n = 1640$) represented 248 experiments from 212 journal articles, VR_{NS} ($n = 80$) represented 13 experiments from 13 journal articles, and VVIQ_{BC} ($n = 62$) represented 7 experiments from 7 journal articles. A 2 (VR, VVIQ) × 2 (BC, NS) study design was used to test the abovementioned research hypotheses (i) effect sizes on VR are larger than those on VVIQ; (ii) effect sizes on NS are larger than those on BC; and (iii) there is an interaction between these two factors.

Results

A preliminary analysis was first performed, wherein a three-level meta-analytic framework was employed to determine the overall effect size estimate (ESE), and parameter estimates for the entire dataset. The average ESE (Z_r) with its 95% Wald confidence interval (CI) was 0.4011 [0.3795, 0.4227], where the level-2 and level-3 heterogeneity variances (τ^2) were 0.0434 and 0.0404, respectively. The test on the null hypothesis of equality of population level-2 and level-3 heterogeneity variances is not statistically significant, χ^2 (1) = 0.4913, $P = 0.4833$. The percentage of variation accounted for at level-2 and level-3 (I^2) were 0.4572 and 0.4249, respectively. This indicates the effect sizes have similar degree of variation within and between studies. Subsequently, a moderator variable was created for each focal category, and the data from each focal category were subjected to an independent three-level meta-analytic framework, such that an internal ESE for each category could be calculated. The ESE and parameter estimates for the four focal categories are presented in Table 1 and Fig. 2. In comparison to the overall model, the effect of the moderator variables was statistically significant, χ^2 (3) = 65.03, $P < 0.001$.

A 2 Vividness (VR, VVIQ) × 2 Measure (BC, NS) design was used to test the research hypotheses with a mixed-effects meta-analysis. The interaction between Vividness and Measure was not statistically significant χ^2 (1) = 1.71, $P = 0.19$, which indicated that the effect of Vividness and Measure was additive. In other words, the effect of Vividness is independent of the effect of Measure (and vice versa). We may independently interpret the effects of Vividness and Measure. When there is an interaction, however, the effect of Vividness depends on the level of Measure (and vice versa). We need to select the level of Measure when we interpret the effect of Vividness (and vice versa). Both Vividness and Measures were significant χ^2 (2) = 63.31, $P < 0.001$, level-3 $R^2 = 0.182$. The ESE for VR was statistically greater than VVIQ ($\Delta Z_r = 0.133$, 95% Wald CI [0.093, 0.173]), after controlling the effect of Measure. NS was statistically greater than BC ($\Delta Z_r = 0.257$, 95% Wald CI [0.150, 0.364]), after controlling the effect of Vividness. Finally, concerning the peripheral analysis examining the direct relationship between VR and VVIQ, the results from 47 effect sizes nested within 19 studies using a three-level meta-analysis suggests an ESE of 0.3977 (0.2886, 0.5069), where the level-2 and level-3 heterogeneity variances (τ^2) were 0.0260 and 0.0293, respectively.

Discussion

The results from the Phase 1 offered significant insight into the ESEs for each of the four focal categories. Although the differences between VR and VVIQ appear to be robust within the BC dimension, the differences within the NS dimension were not as clear cut. Nevertheless, the peripheral category directly correlating VR and VVIQ suggests a weak to moderate relationship. Comparisons concerning the NS dimension did not resolve any differences between VR and VVIQ; however, this effect may have remained unresolved for at least two reasons. First, standard error (SE) within each of the NS focal categories was larger than those within the BC focal categories, which resulted from a much smaller sample size, and a much larger estimate of within study variance ($\bar{v}'$). Secondly, between-study heterogeneity (τ^2) was generally larger within the NS focal categories.

Two strategies were implemented to homogenize the NS focal categories. First, because no future searches indicated additional evidence pertaining to PET ($n = 5$), NIRS ($n = 6$), or pharmacological NS ($n = 4$) neuroscience outcomes, and to avoid overestimating between study error, only fMRI ($n = 73$), and EEG ($n = 54$) were selected as representative to the NS sample. Incidentally, this methodological conceptualization may lend insight into temporal specific (EEG) and spatial specific (fMRI) variations in neurophysiological measurement within this construct. If indicative, a differential propensity for the NS dimension may be observed between fMRI and EEG, a pattern of which should be emergent between both VR and VVIQ. Secondly, another systematic literature search was performed to acquire previously undiscovered journal articles pertaining to VR and VVIQ in the context of NS, but specifically those relating to fMRI and/or EEG.

Table 1. Descriptive statistics for phase 1, including the number of statistical outcomes, number of journal articles, number of experiments, ESEs, and 95% Wald CIs for each category

Parameters	VR_{BC}	$VVIQ_{BC}$	VR_{NS}	$VVIQ_{NS}$
Statistical outcomes	1760	1640	80	62
Journal articles	194	212	13	8
Experiments	238	248	13	8
ESE	0.461	0.323	0.662	0.669
95% Wald CI	[0.432, 0.491]	[0.295, 0. 351]	[0.528, 0.795]	[0.501, 0.837]

Method: Phase 2

Sampling

In an effort to clarify the meaning of the NS data specifically, an exhaustive systematic search was performed, which sought to discover any and all journal articles relating VR and/or VVIQ to EEG and/or fMRI. Four strategies were employed. First, PsychINFO was searched on July 20th, 2015, using the specific parameters “vivid* fMRI”, and “vivid* EEG”, from which six novel journal articles were discovered (Rabin et al. 2010; Summerfield et al. 2010; Todd et al. 2013; Blumen et al. 2014; Johnson et al. 2015; Knyazev et al. 2015) resulting in 37 novel observations (VR_{fMRI}, $n = 30$; VR_{EEG}, $n = 4$; $VVIQ_{fMRI}$, $n = 3$; $VVIQ_{EEG}$, $n = 0$). Similar searches through PsychINFO for VVIQ did not return any results.

Second, the entire library catalog in our institution was searched on July 20th, 2015 using the specific parameters “VVIQ fMRI,” and “VVIQ EEG,” from which three novel journal articles were discovered (Amedi et al. 2005; Robineau et al. 2014; Shen et al. 2015), resulting in 27 novel observations (VR_{fMRI}, $n = 6$; VR_{EEG}, $n = 0$; $VVIQ_{fMRI}$, $n = 0$; $VVIQ_{EEG}$, $n = 21$).

Third, the discussion sections of all relevant papers were read, and references were recorded for comparisons the authors sought to explain their findings. This resulted in an additional two novel journal articles (Gilboa et al. 2004; Mohr et al. 2009), resulting in 29 novel observations (VR_{fMRI}, $n = 28$; VR_{EEG}, $n = 0$; $VVIQ_{fMRI}$, $n = 1$; $VVIQ_{EEG}$, $n = 0$). Finally, a journal article known

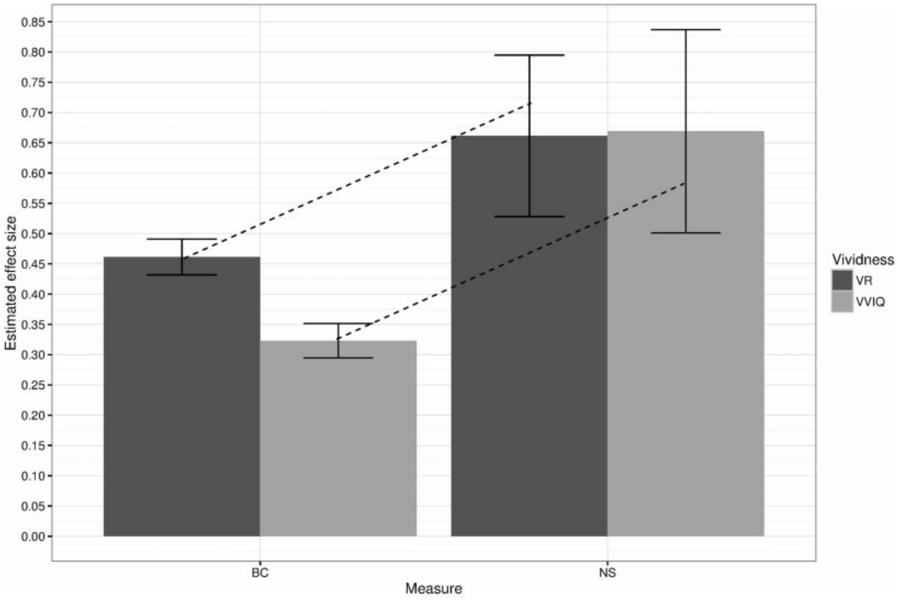

Figure 2. Internal ESEs for each of the four categories in Phase 1, and their 95% Wald CIs. Also plotted are the predicted values from the regression analysis (dotted lines), which are based on the model without the interaction. *Note*: the 95% Wald CIs are based on the estimated values (not the predicted values).

Table 2. Descriptive statistics for phase 2, including the number of statistical outcomes, number of journal articles, number of experiments, ESEs, and 95% Wald CIs for each category

Parameters	$\mathrm{VR}_{\mathrm{BC}}$	$\mathrm{VVIQ}_{\mathrm{BC}}$	$\mathrm{VR}_{\mathrm{EEG}}$	$\mathrm{VVIQ}_{\mathrm{EEG}}$	$\mathrm{VR}_{\mathrm{fMRI}}$	$\mathrm{VVIQ}_{\mathrm{fMRI}}$
Statistical outcomes	1760	1640	35	44	102	43
Journal articles	194	212	6	4	13	8
Experiments	238	248	6	4	13	8
ESE	0.463	0.323	0.619	0.644	0.761	0.616
95% Wald CI	[0.432, 0.494]	[0.294, 0.353]	[0.421, 0.818]	[0.420, 0.867]	[0.629, 0.892]	[0.413, 0.820]

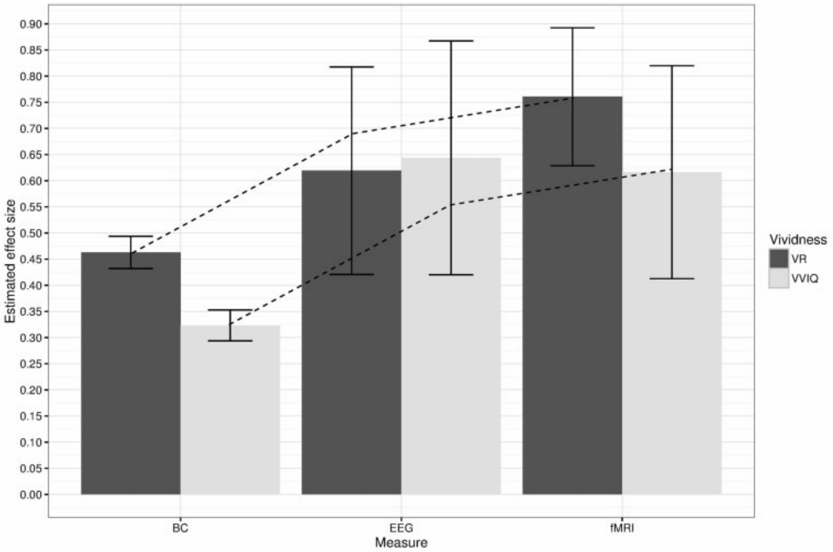

Figure 3. Internal ESEs for each of the six categories in Phase 2, and their 95% Wald CIs. Also plotted are the predicted values from the regression analysis (dotted lines), which are based on the model without the interaction. *Note*: the 95% Wald CIs are based on the estimated values (not the predicted values).

to be relevant (Bird *et al.* 2010), and published in a pilot study through this research group, and not found in any other search was also included, resulting in four novel observations ($\mathrm{VR}_{\mathrm{fMRI}}$, $n = 4$; $\mathrm{VR}_{\mathrm{EEG}}$, $n = 0$; $\mathrm{VVIQ}_{\mathrm{fMRI}}$, $n = 0$; $\mathrm{VVIQ}_{\mathrm{EEG}}$, $n = 0$). Data for Phase 2 are available upon request.

Data trimming and data analysis

The novel fMRI and EEG data from Phase 2 were compiled with existing fMRI and EEG data from Phase 1. Subsequently, the entire NS database consisted of 224 observations. Data for $\mathrm{VR}_{\mathrm{fMRI}}$ ($n = 102$) represented 13 experiments from 13 journal articles, $\mathrm{VVIQ}_{\mathrm{fMRI}}$ ($n = 43$) represented eight experiments from eight journal articles, $\mathrm{VR}_{\mathrm{EEG}}$ ($n = 35$) represented six experiments from six journal articles, and $\mathrm{VVIQ}_{\mathrm{EEG}}$ ($n = 44$) represented four experiments from four journal articles. BC data from Phase 1 were also borrowed from Phase 1, resulting in a 2 (VR, VVIQ) × 3 (BC, EEG, fMRI) study design. Phase 2 data were modeled in the same way as Phase 1. A 2 (VR, VVIQ) × 3 (BC, EEG, fMRI) study design was used to test the abovementioned research hypotheses (i) effect sizes on VR are larger than those on VVIQ; (ii) effect sizes on EEG and fMRI are larger than those on BC; and (iii) there is an interaction between these two factors.

During the interval of time required for peer-review, a final library search was conducted to double check that the database was still representative of the very latest publications. The search revealed that only three new (fMRI) papers published during the peer-review lag (Bergmann *et al.* 2016; Motoyama and Hishitani 2016; Dijkstra *et al.* 2017) fitting the inclusion criteria were not included in our revised analysis. All the data points from these papers were observed to fall within the 95% CIs calculated in Phase 2 (Table 2).

Results

A preliminary analysis was first performed, wherein a three-level meta-analytic framework was employed to determine the overall ESE, and parameter estimates for the entire dataset. The average ESE (Z_r) with its 95% Wald CI was 0.4080 [0.3854, 0.4306], where the level-2 and level-3 heterogeneity variances (τ^2) were 0.0427 and 0.0472, respectively. The test on the null hypothesis of equality of population level-2 and level-3 heterogeneity variances is not statistically significant, χ^2 $(1) = 1.03$, $P = 0.31$. Therefore, the percentage of variation accounted for at level-2 and level-3 (I^2) were 0.4214 and 0.4660, respectively. This indicates the effect sizes have similar degree of variation within and between studies. Subsequently, a moderator variable was created for each focal category, and the data from each focal category were subjected to an independent three-level meta-analytic framework, such that an internal ESE for each category could be calculated. The ESE and parameter estimates for the six focal categories are presented in Table 2 and Fig. 3. In comparison to the overall model, the effect of the moderator variables was statistically significant, χ^2 (5) = 76.77, $P < 0.001$.

A 2 Vividness (VR, VVIQ) × 3 Measure (BC, EEG, fMRI) design was used to test the research hypotheses with a mixed-effects meta-analysis. The interaction between Vividness and Measure was not statistically significant χ^2 (2) = 1.14, $P = 0.57$, which indicated that the effect of Vividness and Measure was additive. In other words, the effect of Vividness is independent of the effect of Measure (and vice versa). Both Vividness and Measures were significant χ^2 (3) = 75.63, $P < 0.001$, level-3 $R^2 = 0.214$. The ESE for VR was statistically greater than VVIQ ($\Delta Z_r = 0.137$, 95% Wald CI [0.095, 0.178]), after controlling the effect of Measure. By using BC as the reference group, EEG was statistically greater than BC ($\Delta Z_r = 0.229$, 95% Wald CI [0.079, 0.379]), and the predicted ESE for fMRI was likewise statistically greater than BC ($\Delta Z_r = 0.297$, 95% Wald CI [0.184, 0.410], after controlling the effect of Vividness. There was no difference between the ESE for EEG and the ESE for fMRI, χ^2 (1) = 0.52, $p = 0.47$.

General Discussion

The present research analyzed a robust and representative subset of literature in an effort to understand the relationship between trial-by-trial VR and VVIQ, within the context of behavioral/cognitive (BC) and neuroscience (NS) experimental paradigms in Phase 1, and within the context of BC, EEG, and fMRI experimental paradigms in Phase 2.

Through the preliminary analysis in Phase 1, a significant difference between the ESEs for VR and VVIQ in the context of BC experimentation was observed. This result is limited in so far as we assume the effect sizes used to create the database have the same statistical meaning, and that there are no systematic differences between the way in which VR and VVIQ are implemented in research. For example, it is possible that VVIQ is more conducive to between-groups comparison, and VR more conducive to within-group comparisons, which may imply the latter has inherently greater effect sizes. It is unlikely, however, that this limitation can account for the differences which were observed across these multiple domains of psychology and neuroscience. For example, the average correlation between VR and VVIQ was generally weak ($r \sim 0.40$), for two measures which purportedly "measure" the same thing [note the correlation reported here is similar to the one reported by Pearson et al. (2011)]. Although differences between VR and VVIQ were observed in Phase 1, the methodology utilized in Phase 2 sought to homogenize, and increase the sample size of the NS dimension, such that differences between the ESEs for VR and VVIQ could be observed within the NS dimension as well.

The statistical methodology employed in Phase 2 was designed to evaluate the data at different levels. Data were first analyzed at an overall level, then analyzed at theoretically determined levels, which were represented by six independent categories belonging to BC, EEG, or fMRI experimental types. The three-level meta-analysis provides a correct statistical model to handle multiple effect sizes nested within studies (Cheung 2014). The ESE, when calculated over the entire dataset was approximately $Zr = 0.40$. Interestingly, the level-2 and level-3 estimates of heterogeneity, or the proportion of variation accounted for at each level (I^2), show a similar magnitude ($\sim$0.42, and $\sim$0.47, respectively). This finding may suggest just as much variation exists within experiments as between. Finally, the unique contributions of Vividness and Measure were tested by a three-level mixed-effects meta-analysis with the data from Phase 2.

The results did not support the interaction between Vividness and Measure. This seems to suggest that the effects of Vividness and Measures are additive. On the other hand, the results provided strong support for both major hypotheses. VR demonstrated larger ESEs than the VVIQ, when compared against a large and representative sample, and larger ESEs were also observed for EEG and fMRI when compared with BC. Although BC experimental paradigms generally result in smaller ESEs, no significant difference was observed between EEG and fMRI neuro-imaging experiment types; albeit, the small sample sizes do not lend to a strong conclusion. Although fMRI has a tendency to overinflate effect sizes in some psychological contexts (Vul et al. 2009), presumably based on statistical power, and limitations arising from small sample sizes (Button et al. 2013), the results of the present analysis do not suggest the ESE for fMRI data deviate significantly from EEG data. However, differences between VR and VVIQ remained unobserved within the EEG dimension, as this level of analysis was presumably limited by very small sample sizes (six and four experiments, respectively).

In relation to other subjective measures used in consciousness research, vividness could be interpreted as the measurement of a specific type of conscious experience, correlate of imagery. Although there have been very few studies comparing imagery vividness to the different types of subjective scales used in consciousness research such as the perceptual awareness scale (PAS, e.g. Ramsøy and Overgaard 2004) or confidence ratings (CR, e.g., Cheesman and Merikle 1986), few older studies (pre-dating PAS creation) which compared VR in perception and imagery seem to show robust relationships (Standing 1973; Giusberti et al. 1992). Relatively more studies showed that VVIQ, and to some extent VR, are correlated with CR [see Baddeley and Andrade (2000), for review; see also Pearson et al. (2011)] in that as people rate themselves as having relatively more vivid images, they are also more confident about their imagery-related ability and performance (and females do so more than males). In addition, in his meta-analysis McKelvie (1995) reported on a subset of 15 studies correlating VVIQ and objective responses and subjective judgments in perceptual tasks, with composite effect size of $r = 0.45$ [95% CI = 0.31, 0.56]. Thus, issues of reportability for subjective measures of awareness used in consciousness research can shed some insights into vividness as a characteristic of conscious experiences in general. As reviewed by Timmermans and Cleeremans (2015), the validity and reliability of subjective measures of conscious awareness can be assessed through a set of reportability criteria: "exhaustiveness" (of all relevant contents in consciousness) and "exclusiveness" (of relevant conscious processes but not unconscious or irrelevant conscious ones), as well as how "directly" or "indirectly" the scales, ratings or reports, measure these attributes in underlying processes as defined by reference to objective measurements. Similar criteria can be applied to VR and VVIQ to better understand vividness as an attribute of visual consciousness. Hence, where relevant in the following discussion, we will highlight points of connection and the reciprocal implications for vividness as type-specific imagery experience measure and as an attribute sharing aspects common to other forms of consciousness as measured by PAS, CR, and other subjective measures.

If we consider VR as an entire set of all phenomenological experiences we can probe (we can ask participants to rate the vividness of mental images generated in any number of single trials), the VVIQ can only approximate this set, as it makes a series of vividness measurements from a small, standardized subset of all possible mental images we can imagine. Albeit, the questions used in the VVIQ may be highly exemplary/representative of the general vividness construct. According to this reasoning, the VVIQ is a robust subset of all VR. Although the 16 or 32 items on the VVIQ/VVIQ2 possess excellent content validity, they cannot accommodate special circumstances in human psychological experience, like those concerning flash-bulb memories in PTSD, or vivid mental images as they relate to mental health more generally (Brewin *et al.* 2010). As such, the greater effect sizes observed for VR observations is likely the surplus of what the VVIQ cannot measure. If VR can be seen as a specific type of subjective awareness measure such as PAS or CR used in consciousness research, then VR can be said to be more "exhaustive" than VVIQ as a measure of the conscious states associated with imagery, in that it reveals most of the relevant knowledge of the conscious state experienced by the imager.

In comparison to BC, larger ESEs were observed for EEG and fMRI experimental paradigms, which may lend weight to the interpretation that, when behavioral and cognitive factors are ruled-out, something extra remains unexplained in the statistical model, yet is captured by neuroimaging studies. Vividness ratings can be captured on a millisecond to millisecond temporal basis (EEG), and in a millimeter to millimeter spatial basis (fMRI), in a way which most closely approximates the conscious "psychological" experience of the mental image. NS experimental paradigms allow for the resolution of specific and intrinsic processes, closely following the human psychological experience of vividness, which is an otherwise unobservable phenomenon. Given this interpretation requires extensive support, Phase 2 was designed to exhaustively retrieve any and all evidence linking vividness to brain modulation. A series of searches afforded 29 relevant papers, representing 224 effect sizes, which were statistically modeled with a series of three-level meta-analyses. The three-level meta-analysis has high statistical power as it includes all effect sizes and models the dependence of the data properly. The significance of the present results is contingent on the weight of a corpus of BC experimentation, which showed a similar pattern, but with a much larger sample size – and thusly – contribution to the overall analysis.

As previously mentioned, one possible limitation to the present analysis may be that VR and VVIQ correspond to different methodological designs, and the inherent lack of heterogeneity between these measures may warrant careful reconsideration of the abovementioned interpretation. However, systematic differences in the way VR and VVIQ are administered, if any, are marginal at best, and cannot account for the present pattern of results. First, the correlation between the VVIQ and VR was weak to moderate. Secondly, the present argument can be thought of as a trade-off in the implementation of the VVIQ, where the convenience of administering a questionnaire (which measures trait ability) sacrifices resolution moment-to-moment. VRs seem to reasonably resolve the spectrum of variability in the mental imagery experience (which can be assumed to include measurement of both trait and state ability). It is this variability which can be more sensitively correlated to modulations in brain activity, which exist within a narrow temporal and/or spatial window, the likes of which a questionnaire cannot impute. Just to clarify with an example, our scores on the VVIQ may very well be the same, but the vividness of events relating to my PTSD (if I suffered from it), would be wildly different than yours (if you did not). Such situational variability may be consistent with flash-bulb memories, pharmaceutical interventions, hallucinations, or dreams. Furthermore, we do acknowledge that the practical interpretability of the average effect size for each main category may be limited (for predictive purposes). However, the relativistic differences (VR > VVIQ) are sufficiently accurate, robust, withstanding replication and convergent validation.

Another potential limitation is that there is a disparity between the number of studies as well as the number of effect sizes in Phase I (VRBC, $n = 1826$; VVIQBC, $n = 1680$; VRNS, $n = 82$; VVIQNS, $n = 62$). This disparity indicates that traditionally BC has been much more studied than NS. The estimated average effects (VR versus VVIQ and BC versus NS) are still unbiased regardless of the disparity of the number of effect sizes. On the other hand, the estimated heterogeneity variances are likely biased toward BC as there are more studies in BC.

There are pros and cons of including unpublished studies in a meta-analysis. Ferguson and Brannick (2012) argued against including unpublished studies. For example, unpublished studies are usually of weaker methodology. Moreover, the search for unpublished studies may also be biased due to the availability of unpublished studies to the authors. On the other hand, Rothstein and Bushman (2012) provided counter arguments for including unpublished studies. They suggested unpublished studies may be excluded on the methodological rigorousness by using clearly defined inclusion criteria rather than excluding all unpublished studies. The present study only included effect sizes from published studies because we wanted to include peer-reviewed studies with rigorous methodology. However, we do not think that the presence of unpublished studies may alter our conclusions because there are more than 3600 effect sizes nested within about 400 studies. It is not feasible to find enough unpublished studies with null effect substantially changing our conclusions.

If one considers the totality of the available data in comparison to theoretically established categories, cluster/types of research can be independently identified, and decomposed from the overall analysis. The entire BC dimension could, by necessity, be further reduced into cluster/types (as were EEG and fMRI data decomposed from the Phase 1 NS dimension), which can then be objectively compared, contrasted, and evaluated. In fact, the 32-month delay between the searches in both Phases represents a temporal necessity, allowing the collection of an incremental amount of literature pertaining to the spatial/temporal imaging of vividness, which remains a specialist field of ongoing research debate. After this interval of time, enough literature was compiled to reflect trends previously observed within the BC dimension. The replication of low "overall" effect sizes throughout the BC subdimensions is a consistent finding; here, ΔZr ($\sim$0.14) is estimated to be the general differential throughout every possible subset comparison of VR and VVIQ within the BC dimension. The estimation is supported at an overall level, theoretically relevant NS levels, and replicated in two previous smaller subset analyses (D'Angiulli *et al.* 2013; Runge *et al.* 2015).

The question is, then, at what "level" does one want to cluster the BC subsets? At some point, researchers are required to make a conceptual distinction between two or more types of research within an overall dataset. Indeed, "Behavioural" experimentation compared to "Cognitive" experimentation may be the simplest BC comparison one could make, the effect of which was already measured as 0.40 "overall" – the VRs $\sim$ 0.46, and the VVIQs $\sim$ 0.32. Conversely, if one wants to make absolutely no assumptions concerning heterogeneity, only direct correlations between VR and VVIQ can be considered valid comparisons (because they are within studies, and unambiguously summarize the relationship), the effect size of which is similarly ~0.40. Making relevant theoretical distinctions [qualified by theory, quantified by estimates of heterogeneity (τ^2)] enables meta-analysis. For example, by concentrating research efforts on peer-reviewed quality reports and filtering out the gray literature, an acceptable level of journal quality heterogeneity was assumed, which was defined by the requirement of empirically evaluated, evidenced-based research (Hopewell *et al.* 2005).

Fundamentally, the consistent observation that ESEs for VR exceed those of VVIQ suggests convergent validation, and under such circumstances, it is reasonable to conclude that VR is a more reliable self-report measure than VVIQ. In this context, however, we can go a step further and define "more reliable" concretely in terms of underlying processes, which are related to reportability. That is, the present findings support the conclusion that VR offer a relatively more direct report of imagery, in that they may be more sensitive to immediate, unfiltered information about the visual qualities of mental images relevant to a task or a resting condition, and may show knowledge that is relatively more exclusive to sensory-perceptual consciousness. In contrast, VVIQ is more indirect, since it could include higher-order and metacognitive processes related to the self-judged expected ability to generate at will a visual image that corresponds to a complex scene (which includes relatively more abstract knowledge, and needs to be translated from a fairly complex verbal narrative). In many respects, the difference between VR and VVIQ parallels the difference between PAS and CR in consciousness research, where the former measure has been shown to be correlated more with perception than metacognition, whereas the latter shows the opposite pattern of correlation (Timmermans and Cleeremans 2015). All the pros and cons of using PAS and CR, as those authors have well pointed out, also apply for VR and VVIQ. For example, both vividness measures might fail in being exclusive, because they might reflect information below the level the experimenter is interested in, or because unconscious processes may exert an influence by boosting or attenuating the ratings. However, it is unclear whether a "reverse subtractive logic" can be pursued in comparing proportions of accounted variation in the data between VR and VVIQ, as it could be done for PAS and other subjective awareness measures. Could we infer that the differential "unexplained" variance between VR and VVIQ indicates underlying overflow conscious processes plus underlying unconscious processes? This is an interesting empirical question for future research.

The operational partition in neural and behavioural processes is a feature derived from the way the field has historically developed, and it reflects a legitimate reduction of the investigative approach: it is not practical, nor scientifically necessary to consider all possible variables and measures. However, it still remains that such partition is artificial. Phenomenology, neural, cognitive and behavioral processes are all components of an integrated system. The finding that both VR and VVIQ are more strongly associated with the neural components of imagery generation than the cognitive and behavioral ones implies that if neural measures are considered the third independent criterion of reference for validity, then VR, regardless of the route of reportability they implement – VR or VVIQ – may in some circumstances be more valid than behavioral (and cognitive) measures. The findings and theoretical implications of the present meta-analysis lend support to the view that typical findings in the literature which show low correlations between self-report and behavioral measures should not be necessarily interpreted as a weakness of the former, rather the latter. Studies that do not adopt a single behavioral measure as the gold standard and criterion variable, and use multiple formats of the construct of interest, show that construct validity coefficients of self-reports such as VR are invariably greater than their behavioral counterparts [see review in Haeffel and Howard (2010)].As earlier noted by Jack and Roepstorff (2003), self-reports are an aspect of consciousness that needs to be established with brain patterns, to explain the nature of experience. Inevitably, they need to be understood. However, our "approximation" of what the sentient unit is experiencing at any particular moment in time grows as we fumble with reductionism, and deconstruction, consistent with an information-processing approach. Validity of vividness seems to be proportional to the correlation between vividness consciously experienced and vividness verbally reported, and it is a misjudgment to assume we cannot trust subject's verbal interpretations of their own conscious states. In other words, it is a misjudgment to assume that validity needs to be established only by robust correlations with behavioral measures.

The previous conclusion has all but a trivial implication since the data we have presented clearly do not support current approaches which call for "eliminating" subjective measures, thereby only focusing on the relationship between behavioral and neural measures of conscious processes [see again Timmermans and Cleeremans (2015)]. Expanding on the Jack and Roepfstorff's triangulation challenge, Goldman (2004, p. 9) offered a compelling argument that from a possible lack of correlation between vividness and behavioral accuracy (for instance in memory), it is not possible to refute the correspondence between vividness measures and actual experienced vividness. The accuracy of the vividness measures may still be valid even if there is no correspondence between behavioral accuracy and experienced vividness. Our study contributes to this specific argument in showing that, empirically, brain activity measures correspond closely to vividness measures, and to a lesser degree to cognitive and behavioral measures. Consequently, matters appear to be a bit more complex: it is not so much about the absence of correspondence, since across many experiments all the terms in the set of relationships are to some extent reciprocally correlated. It is about the strength of the correlations, i.e., empirically what matters is the degree of precision (reliability) for which some relationships stand out more than others.

Vividness and Consciousness: Theoretical Underpinnings of Validity

Finally, we would like to discuss some implications of a more general, less technical nature concerning how this study is situated to advance our understanding of vividness and consciousness in general. That is, in terms of the validity approach we have taken this is the next logical step: we attempt here to provide a full blown (albeit concise) theory of the structure of the attribute which the term vividness refers to.

The present study is compatible with a tenet that is widely held in some current neurobiological theories of consciousness. In brief, that sensory-affective (not exclusively/necessarily visual) mental images, defined as objective "isomorphic neural maps" associated with external and inner (somatic) environmental input, can constitute the first-order basic building block of consciousness during both perception and memory of perception [for review see Feinberg and Mallatt (2016)]. Within this broader theoretical framework, a corollary that can be added through this study is that vividness expresses the graded isomorphic aspect of the conscious experience that more or less directly and immediately arises from the brain's complex processing of sensory-affective information in perception and memory. That is, vividness could be conceived as a particular isomorphism or equivalence mapping function (Levesque 1986) that goes beyond the "turned on" consciousness state; as it expresses changes and states of inner experience in terms of equivalent graded levels of knowledge. In this context, vividness may have the role of making explicit the level or strength of the isomorphic correspondence between the inner first-person experience of the input and the environmental input itself (Marks 1999). In other words, measurements of vividness, notably through VR, reflect the gradient of immediate conscious experience of images that mirror isomorphic objective neural processes during perception and memory. Thus, vividness can be considered a chief phenomenological feature of primary sensory consciousness, and it supports the idea that consciousness is a graded phenomenon.

The concept of vividness however can be extended to a number of other types of consciousness. Isomorphic maps may develop into and become embedded in more complex and dynamic neurofunctional structures ['nested neural hierarchies' Feinberg (2011); "activity cycles" (Freeman, 1983)] that include emotions, plans, goals and actions, etc., which are removed from immediate environmental inputs or memory of them, in other words, states of consciousness other than those associated with perception and memory. These states may involve derivative "second-order" isomorphic correspondences (Shepard 1987) whose organization is not in forms of maps, but complex multidimensional patterns [e.g. the computational conceptual structures described by Tononi *et al.* (2016)]. In this other context, vividness may be recycled to express graded knowledge of conscious experience that may or may not be linked with any first-order isomorphic representations, and it may or may not recycle sensory-affective isomorphic maps in memory. In these cases, vividness may express conscious knowledge even though it is not necessarily experienced in the form of mental images. Hence, vividness may have also an important role in accounting for phenomenal conscious processes associated with mental simulation (Marks 1999), analogical reasoning (Levesque 1988), and strategies and metacognition (Hertzog and Dunlosky 2006). As suggested by our findings, VVIQ may be a particularly sensitive measure of vividness when the latter cases involve imagery. However, other subjective or indirect computational measures could be devised to indicate how vividness captures other aspects of conscious knowledge which do not involve or require mental images at all (Brachman and Levesque 2004). Indeed, as mentioned in the introduction, we believe one of the most promising potential future contributions of the construct of vividness may be allowing a constructive integration of the roles of emotions, motives and goals within neurocognitive and epigenetic theories of imagery and consciousness.

Conclusion

Although the first attempt to validate the vividness construct through meta-analysis can be attributed to McKelvie (1995), his detailed and systematic study was limited to the VVIQ. The VVIQ became the most widely used, standard tool by which vividness, and imagery self-reports are studied (Pearson 1995), but this conclusion may be predicated on an erroneous assumption. The present analysis suggests the study of vividness, and researcher conceptualization of it more generally, needs to be reconsidered. In addition, it suggests that the construct of vividness may be most thoroughly studied using neuroscience methodologies that do not necessarily have observable behavioural outputs, such that phenomenological self-reports may be reliably and validly associated with neural correlates on a trial-by-trial basis. Capitalizing on the theoretical and methodological underpinnings of its validity, the concept of vividness can explain key aspects of the phenomenological experience of mental imagery, but it can be applied beyond, extending to other forms of conscious awareness, which do not necessarily involve imagery.

Acknowledgements

This project was supported by a standard research grant to Amedeo D'Angiulli awarded from the Social Sciences and Humanities Research Council of Canada, grant no. 410-2011-1463.

Supplementary data

Supplementary data is available at Neuroscience of Consciousness *Journal* online.

Conflict of interest statement. None declared.

References

Alter AL, Balcetis E. Fondness makes the distance grow shorter: desired locations seem closer because they seem more vivid. *J Exp Soc Psychol* 2010;**47**:16–21.

Amedi A, Malach R, Pascual-Leone A. Negative BOLD differentiates visual imagery and perception. *Neuron* 2005;**48**:859–72.

Andrade J (Ed.). *Working Memory in Perspective*. Hove, East Sussex, England: Psychology Press, 2001.

Andrews-Hanna JR, Reidler JS, Sepulcre J, *et al*. Functional-anatomic fractionation of the brain's default network. *Neuron* 2010;**65**:550–62.

Baddeley AD, Andrade J. Working memory and the vividness of imagery. *J Exp Psychol Gen* 2000;**129**:126–45.

Barrowcliff AL, Gray NS, Free-Man TCA, *et al*. Eye-movements reduce the vividness, emotional valence and electrodermal arousal associated with negative autobiographical memories. *J Forens Psychiatry Psychol* 2004;**15**:325–45.

Bergmann J, Genç E, Kohler A, *et al*. Smaller primary visual cortex is associated with stronger, but less precise mental imagery. *Cerebral Cortex* 2016;**26**:3838–50.

Betts GH. *The Distribution and Functions of Mental Imagery (No. 26)*. New York, NY: Teachers College, Columbia University, 1909.

Bird CM, Capponi C, King JA, *et al*. Establishing the boundaries: the hippocampal contribution to imagining scenes. *J Neurosci* 2010;**30**:11688–95.

Blumen HM, Holtzer R, Brown LL, *et al*. Behavioral and neural correlates of imagined walking and walking-while-talking in the elderly. *Hum Brain Map* 2014;**35**:4090–104.

Borenstein M, Hedges LV, Higgins J, *et al*. *Introduction to Meta-Analysis*. Hoboken, NJ: John Wiley & Sons, Ltd, 2009, 149–86.

Borsboom D. *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press, 2005.

Borsboom D, Mellenbergh GJ, van Heerden J. The concept of validity. *Psychol Rev* 2004;**111**:1061–71.

Brachman RJ, Levesque HJ, (2004). *Knowledge Representation and Reasoning*. Massachusetts, USA: Morgan Kaufmann Publishers.

Brewin CR, Gregory JD, Lipton M, *et al*. Intrusive images in psychological disorders: characteristics, neural mechanisms, and treatment implications. *Psychol Rev* 2010;**117**:210.

Button KS, Ioannidis JP, Mokrysz C, *et al*. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;**14**:365–76.

Bywaters M, Andrade J, Turpin G. Determinants of the vividness of visual imagery: the effects of delayed recall, stimulus affect and individual differences. *Memory* 2004;**12**:479–88.

Campos A, Amor A, Gonzalez MA. Presentation of keywords by means of interactive drawings. *Spanish J Psychol* 2002;**5**:102–9.

Cheesman J, Merikle PM. Distinguishing conscious from unconscious perceptual processes. *Can J Psychol* 1986;**40**:343–67.

Cheung MWL. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol Methods* 2014;**19**:211.

Cheung MWL. metaSEM: an R package for meta-analysis using structural equation modeling. *Front Psychol* 2015;**5**:1521.

Cichy RM, Heinzle J, Haynes JD. Imagery and perception share cortical representations of content and location. *Cereb Cortex* 2012;**22**:372–80.

Conduit R, Crewther SG, Coleman GJ. Spontaneous eyelid movements (ELMS) during sleep are related to dream recall on awakening. *J Sleep Res* 2004;**13**:137–44.

Cooper, H, Hedges, LV, Valentine, JC (eds.), *The Handbook of Research Synthesis and Meta-Analysis*. New York, NY: Russell Sage Foundation, 2009.

Cui X, Jeter CB, Yang D, *et al*. Vividness of mental imagery: individual variability can be measured objectively. *Vision Res* 2007;**47**:474–8.

D'Angiulli A. Mental image generation and the contrast sensitivity function. *Cognition* 2002;**85**:B11–9.

D'Angiulli A. Is the spotlight an obsolete metaphor of "seeing with the mind's eye"? A constructive naturalistic approach to the inspection of visual mental images. *Imagin Cogn Pers* 2008;**28**:117–35.

D'Angiulli A, Reeves A. Generating visual mental images: latency and vividness are inversely related. *Memory Cogn* 2002;**30**:1179–88.

D'Angiulli A, Reeves A. Probing vividness increment through imagery interruption. *Imagin Cogn Pers* 2003–2004;**23**:79–88.

D'Angiulli A, Reeves A. The relationship between self-reported vividness and latency during mental size scaling of everyday items: phenomenological evidence of different types of imagery. *Am J Psychol* 2007;**120**:521–51.

D'Angiulli A, Runge M, Faulkner A, *et al*. Vividness of visual imagery and incidental recall of verbal cues, when phenomenological availability reflects long-term memory accessibility. *Front Psychol* 2013;**4**:1.

Dean GM, Morris PE. The relationship between self-reports of imagery and spatial ability. *Br J Psychol* 2003;**94**:245–73.

Dehaene S. *Consciousness and the Brain: Deciphering How the Brain Codes our Thoughts*. New York, NY: Penguin, 2014.

Dijkstra N, Bosch S, van Gerven MA. Vividness of visual imagery depends on the neural overlap with perception in visual areas. *J Neurosci* 2017;**37**:3022–16.

Ericsson KA, Simon HA. Verbal reports as data. *Psychol Rev* 1980;**87**:215–51.

Ericsson KA, Simon HA. *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: The MIT Press, 1993.

Feinberg TE. The nested neural hierarchy and the self. *Conscious Cogn* 2011;**20**:4–15.

Feinberg TE, Mallatt JM. *The Ancient Origins of Consciousness: How the Brain Created Experience*. Cambridge, MA: MIT Press, 2016.

Ferguson CJ, Brannick MT. Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychol Methods* 2012;**17**:120–8.

Flanagan O, Dryden D. Consciousness and the mind: contributions from philosophy, neuroscience, and psychology. In: *Methods, Models and Conceptual Issues: An Invitation to Cognitive Science*. Cambridge, MA: MIT Press, 1998, 133–72.

Freeman WJ. The physiological basis of mental images. *Biol Psychiatry* 1983;**18**:1107–25.

Galton F. Statistics of mental imagery. *Mind* 1880;**5**:301–18.

Galton F. *Inquiries into the Human Faculty & its Development*. New York, NY: MacMillan Co., 1883.

Georgalis N. *The Primacy of the Subjective: Foundations for a Unified Theory of Mind and Language*. Cambridge, MA: MIT Press, 2006.

Gilboa A, Winocur G, Grady CL, *et al*. Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. *Cerebral Cortex* 2004;**14**:1214–25.

Giusberti F, Cornoldi C, Beni R, *et al*. Differences in vividness ratings of perceived and imagined patterns. *Br J Psychol* 1992;**83**:533–47.

Goldman A. Epistemology and the evidential status of introspective reports I. *J Conscious Stud* 2004;**11**:1–16.

Gonsalves B, Reber PJ, Gitelman DR, *et al*. Neural evidence that vivid imagining can lead to false remembering. *Psychol Sci* 2004;**15**:655–60.

Haeffel GJ, Howard GS. Self-report: psychology's four-letter word. *Am J Psychol* 2010;**123**:181–8.

Hale SM, Simpson HM. Effects of eye movements on the rate of discovery and the vividness of visual images. *Percept Psychophys* 1971;**9**:242–6.

Hebb DO. Concerning imagery. *Psychol Rev* 1968;**75**:466.

Hedges LV, Pigott TD. The power of statistical tests in meta-analysis. *Psychol Methods* 2001;**6**:203.

Hertzog C, Dunlosky J. Using visual imagery as a mnemonic for verbal associative learning: developmental and individual differences. In: Vecchi T and Bottini G (eds.), *Imagery and Spatial Cognition: Methods, Models and Cognitive Assessment*. Amsterdam and Philadelphia, Netherlands/USA: John Benjamins Publishers, 2006.

Hopewell S, Clarke M, Mallett S. Grey literature and systematic reviews. In: Rothstein HR, Sutton AJ, Borenstein M (eds.), *Publication Bias in Meta-Analysis. Prevention, Assessment and Adjustments*. West Sussex, UK: Wiley, 2005, 49–72.

Hox JJ, Moerbeek M, van de Schoot R. *Multilevel Analysis: Techniques and Applications*. New York, NY: Routledge, 2010.

Jack AI, Roepstorff A. Why trust the subject? *J Conscious Stud* 2003;**10**:v–xx.

Johnson MK, Kuhl BA, Mitchell KJ, *et al*. Age-related differences in the neural basis of the subjective vividness of memories: evidence from multivoxel pattern classification. *Cogn Affect Behav Neurosci* 2015;**15**:644–61.

Kensinger EA, Addis DR, Atapattu RK. Amygdala activity at encoding corresponds with memory vividness and with memory for select episodic details. *Neuropsychologia* 2011;**49**:663–73.

Knyazev GG, Savostyanov AN, Bocharov AV, *et al*. Oscillatory correlates of autobiographical memory. *Int J Psychophysiol* 2015;**95**:322–32.

Kosslyn SM. *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press, 1994.

Kosslyn SM, Alper SN. On the pictorial properties of visual images: effects of image size on memory for words. *Can J Psychol* 1977;**31**:32–40.

Kosslyn SM, Sukel KE, Bly BM. Squinting with the mind's eye: effects of stimulus resolution on imaginal and perceptual comparisons. *Memory Cogn* 1999;**27**:276–87.

Lee SH, Kravitz DJ, Baker CI. Disentangling visual imagery and perception of real-world objects. *Neuroimage* 2012;**59**:4064–73.

Levesque HJ. Making believers out of computers. *Artificial Intelligence* 1986;**30**:81–108.

Levesque HJ. Logic and the complexity of reasoning. *J Philosophical Logic* 1988;**17**:355–89.

Logie RH, Pernet CR, Buonocore A, *et al*. Low and high imagers activate networks differentially in mental rotation. *Neuropsychologia* 2011;**49**:3071–7.

Marks DF. Visual imagery differences in the recall of pictures. *Br J Psychol* 1973;**64**:17–24.

Marks DF. New directions for mental imagery research. *J Ment Imagery* 1995;**19**:153–67.

Marks DF. Consciousness, mental imagery and action. *Br J Psychol* 1999;**90**:567–85.

Markovic J, Anderson AK, Todd RM. Tuning to the significant: neural and genetic processes underlying affective enhancement of visual perception and memory. *Behav Brain Res* 2014;**259**:229–41.

Markus KA, Borsboom D. *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning*. New York, NY: Routledge, 2013.

McKelvie SJ. The Vividness of Visual Imagery Questionnaire as a predictor of facial recognition memory performance. *Br J Psychol* 1994;**85**:93–104.

McKelvie SJ. *Vividness of Visual Imagery: Measurement, Nature, Function & Dynamics*. New York, NY: Brandon House, 1995.

Mohr HM, Linder NS, Linden DE, *et al*. Orientation-specific adaptation to mentally generated lines in human visual cortex. *Neuroimage* 2009;**47**:384–91.

Moscovitch M, Rosenbaum RS, Gilboa A, *et al*. Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *J Anatomy* 2005;**207**:35–66.

Motoyama H, Hishitani S. The brain mechanism that reduces the vividness of negative imagery. *Conscious Cogn* 2016;**39**:59–69.

Naccache L. Visual consciousness: an updated neurological tour. In: *The Neurology of Consciousness*. Oxford: Academic Press, 2008, 271–81.

Overgaard M. Introspection in science. *Conscious Cogn* 2006;**15**:629–33.

Pearson DG. The VVIQ and cognitive models of imagery: future directions for research. *J Ment Imagery* 1995;**19**:167–70.

Pearson J, Rademaker RL, Tong F. Evaluating the mind's eye: the metacognition of visual imagery. *Psychol Sci* 2011;**22**:1535–42.

Pylyshyn ZW. *Seeing and Visualizing: It's not What You Think*. Boston, MA: MIT Press, 2003.

Rabin JS, Gilboa A, Stuss DT, *et al*. Common and unique neural correlates of autobiographical memory and theory of mind. *J Cogn Neurosci* 2010;**22**:1095–111.

Ramsøy TZ, Overgaard M. Introspection and subliminal perception. *Phenomenol Cogn Sci* 2004;**3**:1–23.

Reisberg D. The nonambiguity of mental images. In: Cornoldi C, Logie RH, Brandimonte MA, Kaufmann G, Reisberg D (Eds.), *Stretching the Imagination: Represent and Transformation in Mental Imagery*. New York: Oxford University Press, 1996;119–72.

Richardson JT. Gender differences in mental rotation. *Percept Mot Skills* 1994;**78**:435–48.

Robineau F, Rieger SW, Mermoud C, *et al*. Self-regulation of interhemispheric visual cortex balance through real-time fMRI neurofeedback training. *NeuroImage* 2014;**100**:1–14.

Rodway P, Gillies K, Schepman A. Vivid imagers are better at detecting salient changes. *J Individ Dif* 2006;**27**:218–28.

Rothstein HR, Bushman BJ. Publication bias in psychological science: comment on Ferguson and Brannick (2012). *Psychol Methods* 2012;**17**:129–36.

Runge M, Bakhilau V, Omer F, *et al*. Trial-by-trial vividness self-reports Versus VVIQ a meta-analytic comparison of behavioral, cognitive and neurological correlations. *Imagin Cogn Pers* 2015;**35**:137–65.

Shen ZY, Tsai YT, Lee CL. Joint influence of metaphor familiarity and mental imagery ability on action metaphor comprehension: an event-related potential study. *Lang Ling* 2015;**16**:615–37.

Shepard RN. Toward a universal law of generalization for psychological science. *Science* 1987;**237**:1317–23.

Smeets MAM, Dijs MW, Pervan I, *et al*. Time-course of eye movement-related decrease in vividness and emotionality of unpleasant autobiographical memories. *Memory* 2012;**20**:346–57.

Standing L. Learning 10000 pictures. *Quart J Exp Psychol* 1973;**25**:207–22.

St-Laurent M, Abdi H, Buchsbaum BR. Distributed patterns of reactivation predict vividness of recollection. *J Cogn Neurosci* 2015;**27**:2000–18.

Summerfield JJ, Hassabis D, Maguire EA. Differential engagement of brain regions within a 'core' network during scene construction. *Neuropsychologia* 2010;**48**:1501–9.

Timmermans B, Cleeremans A. How can we measure awareness? An overview of current methods. In Overgaard M (Ed.), *Behavioural Methods in Consciousness Research*. Oxford: Oxford University Press, 2015;21–46.

Todd RM, Schmitz TW, Susskind J, *et al*. Shared neural substrates of emotionally enhanced perceptual and mnemonic vividness. *Front Behav Neurosci* 2013;**7**

Tononi G, Boly M, Massimini M, *et al*. Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci* 2016;**17**:450–61.

Van den Noortgate W, López-López JA, Marín-Martínez F, *et al*. Three-level meta-analysis of dependent effect sizes. *Behav Res Methods* 2013;**45**:576–94.

Vul E, Harris C, Winkielman P, *et al*. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect Psychol Sci* 2009;**4**:274–90.

Weiskrantz L, Barbur JL, Sahraie A. Parameters affecting conscious versus unconscious visual discrimination with damage to the visual cortex (V1). *Proc Natl Acad Sci* 1995;**92**:6122–6.

Wheeler ME, Petersen SE, Buckner RL. Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci* 2000;**97**:11125–9.