

## Kinds of access: different methods for report reveal different kinds of metacognitive access

Morten Overgaard and Kristian Sandberg

*Phil. Trans. R. Soc. B* 2012 **367**, 1287-1296

doi: 10.1098/rstb.2011.0425

---

### References

[This article cites 42 articles, 6 of which can be accessed free](#)

<http://rstb.royalsocietypublishing.org/content/367/1594/1287.full.html#ref-list-1>

[Article cited in:](#)

<http://rstb.royalsocietypublishing.org/content/367/1594/1287.full.html#related-urls>

### Subject collections

Articles on similar topics can be found in the following collections

[behaviour](#) (344 articles)

[cognition](#) (208 articles)

[neuroscience](#) (272 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

*Research*

# Kinds of access: different methods for report reveal different kinds of metacognitive access

Morten Overgaard<sup>1,2,\*</sup> and Kristian Sandberg<sup>2</sup>

<sup>1</sup>CNRU, Department of Communication and Psychology, Aalborg University, Kroghstræde 3, 9220 Aalborg Øst, Aalborg, Denmark

<sup>2</sup>CNRU, Hammel Neurorehabilitation and Research Center, CFIN, MindLab, Aarhus University, Noerrebrogade 44, Building 10G, 8000 Aarhus C, Denmark

In experimental investigations of consciousness, participants are asked to reflect upon their own experiences by issuing reports about them in different ways. For this reason, a participant needs some access to the content of her own conscious experience in order to report. In such experiments, the reports typically consist of some variety of ratings of confidence or direct descriptions of one's own experiences. Whereas different methods of reporting are typically used interchangeably, recent experiments indicate that different results are obtained with different kinds of reporting. We argue that there is not only a theoretical, but also an empirical difference between different methods of reporting. We hypothesize that differences in the sensitivity of different scales may reveal that different types of access are used to issue direct reports about experiences and metacognitive reports about the classification process.

**Keywords:** metacognition; consciousness; awareness; subliminal perception; vision

## 1. INTRODUCTION AND DEFINITIONS

Although 'metacognition', 'verbal reports' and even 'introspection' have become legitimate concepts in cognitive neuroscience, they are rarely clearly defined, and their relations to cognition and consciousness are often even more elusive. Here, we attempt to show how these concepts may be understood, how they relate to each other and present empirical arguments, which support a hypothesis that introspection may be a unique type of metacognition, i.e. that introspective reports are empirically different from other kinds of metacognitive reports.

In consciousness research, authors typically refer to their behavioural measures of conscious experience as introspective or metacognitive [1–3], and the two terms are often used interchangeably although they differ from a theoretical perspective. In principle, metacognition is any cognitive process about a different cognitive process, whereas introspection is closely tied to conscious experience. Metacognition is thus a higher-order process with cognition as its object, whereas introspection is a higher-order process with conscious experience as its object.

Although consciousness and cognition by several accounts may be related to highly overlapping brain processes, the concepts are defined differently. Cognition, as described in Neisser's [4] landmark publication

*Cognitive Psychology*, is defined as the transformation and use of sensory input, and how these processes work even in the absence of input, as in hallucinations or imagination. The 'transformations' or 'processes' are not themselves observable, but are inferred based on observations of behaviour. Cognition is thus some sort of processing, and it can usually be functionally defined—a cognitive scientist can examine the workings and purpose of the memory system, for instance. In cognitive science, it is not unusual to investigate certain cognitive states that are said to be about other cognitive states rather than external phenomena—the so-called metacognitive states [5]. An important aspect of cognitive science theory is the idea that our knowledge of ourselves is basically inferential and not based on privileged access to one's own mental processes. Nevertheless, this inferential information, it is argued, is frequently causally related to our actions. It is because we have certain ideas about our own qualities or disadvantages that we decide to take a certain career path or give up on another. It is because we think we would like a particular kind of music that we chose to pay money to attend one particular concert. This way, a person can, in principle, perform a metacognitive judgement without trying to evaluate his mental processes or strategies directly, although at other times, a metacognitive judgement might involve some kind of evaluation of an internal process. This could, for instance, be the case when a participant in a psychological experiment is asked to rate his confidence in having performed a visual discrimination task successfully. In other words, metacognition does not require that the information

\* Author for correspondence ([mortover@rm.dk](mailto:mortover@rm.dk)).

One contribution of 13 to a Theme Issue 'Metacognition: computation, neurobiology and function'.

used to evaluate a mental process is obtained by probing the system (using another, internal mental process), although in some cases it may.

There are several meanings of the word ‘consciousness’. One use of ‘conscious’ is applied to a person’s total or background state (what is sometimes called ‘state consciousness’). A person is conscious, in this sense, if he or she is, say, alert and awake rather than being asleep or even in a coma. However, most attempts to capture the meaning of the term consciousness seem to focus on a second aspect, the contents of consciousness. Some philosophers use terms such as ‘qualia’, whereas others prefer ‘subjective experience’, ‘phenomenal consciousness’, ‘conscious awareness’ and the like.

Consciousness has become a fast-growing topic of interest in empirical research, and consequently scientists are methodologically dependent on more than the presence or absence of consciousness in order to study it. They need participants to be able to give some sort of report or externally observable indication that they experience something, thus leading to a renewed interest in how participants report and introspect. Introspection, however, is hardly a new method. At the beginning of experimental psychology, William James, for one, believed that one method to study the mind should be the direct, inner observation of consciousness [6]. Common to classical as well as more recent discussions of introspection is the necessary link to conscious experience. A broadly discussed account, mentioned by William James among others, defines introspection as a self-reflective, retrospective process in which one’s memory trace of the original target mental state is inspected. In more recent accounts, introspection is understood and investigated as an ‘on-line’ inspection of current and ongoing mental states. The available definitions of introspection agree, however, that it should be defined as some sort of observation of or attention to what is subjectively experienced [6–8], for which reason there cannot be introspection without experience as a matter of definition.

As we can see from the above context, there are many intuitive similarities between the concepts of metacognition and introspection, yet the two are differently defined. Whereas metacognition is functionally defined as basically any cognitive state that is about another cognitive state, introspection can only be about a specifically conscious state. In this sense, both concepts can be said to be of ‘second order’ as they are about another (first-order) state.

On the basis of reasoning so far, the distinctions shown in table 1 can be made. Table 1 shows how introspection and metacognition are defined in different ways although their use overlaps greatly. Metacognition is a concept with a functional definition, whereas introspection is defined from a subjective perspective. This distinction has little to say about the relationship between metacognition and introspection. Clearly, the relationship need not be a matter of opposition. Rather, introspection might be a special kind of metacognition.

#### (a) *Methodological consequences*

Just as consciousness should not be confused with introspection, introspection should not be confused with a

Table 1. Structure of definitions.

	‘first order’ mental state	‘second order’ mental state
functionally defined	cognition	metacognition
subjectively defined	consciousness	introspection

report. A report is, in this context, an intended communication by a participant and may be delivered verbally or by any other kind of controlled behaviour (signs, button presses or whatever). We may have full metacognitive or introspective knowledge of some mental event but choose not to report it, to lie about it or to report just one aspect of it, while not mentioning another. Thus, introspection and metacognition are fully dissociable from reports. However, the opposite is not the case: reports about conscious or cognitive states are not dissociable from introspection or metacognition.

These conceptual distinctions lead logically to a number of methodological consequences. First of all, one methodological consequence is that any neuroscientific investigation of cognitive processes using metacognition or of consciousness using introspection will have difficulties in sorting out which brain activations are related to the first-order and to the second-order states. This is, however, not to say that nothing can be done experimentally to tease the levels apart. For one thing, the second-order states, as a matter of principle, cannot exist without the simultaneous presence of a first-order state, which it is about. First-order states, however, do not stand in any dependent relation to the second-order states. One might assume a certain temporal relation between them, so that the first-order mental states always occur before the second-order state that is about the first-order state.

The number of methodological arguments against the use of introspection to study consciousness is impressive. Most of them, however, circle around one central claim that introspection does not give reliable and valid information about conscious processes. Nisbett & Wilson [9] presented empirical evidence that subjects have little introspective access to the causes of their choices and behaviour. In one example, they showed that participants had a bias to prefer objects presented to the right, yet when asked, they never mentioned the location of the object as their reason for preferring it. However, participants giving an introspective report about liking objects presented to the right for some other reason than the object’s location in space may be giving a perfectly good and scientifically usable report of what they experienced. Nisbett and Wilson correctly rejected introspection as a methodology to learn about (some aspects of) choice and decision-making, as the behavioural data suggested a very different explanation from the one that participants reported. Considering table 1, it should be clear that an introspective report by definition is only about what is experienced and not about cognitive processes [9,10].

The discussion rests on a philosophical debate about whether introspection is corrigible or not, or whether we have ‘privileged access’ to our own

consciousness. In the view presented here, where consciousness is considered subjective and introspection is an attending to consciousness, introspection is incorrigible in the sense that no external measure can directly access a person's experiences. The introspective report, of course, may be corrected by the subject himself on closer examination or if he was lying or otherwise somehow reporting suboptimally the first time. In this sense, you could be asked to calculate, say, 956 minus 394 and come up with the right answer, but when asked how you did it, produce a report describing a method that logically would lead to a wrong answer. This would make your report false as a description of how you actually performed the math. However, it may still be correct as an introspective report, telling what you experience when inspecting your memory of what happened.

Another argument against introspection has been that introspection is not exclusive, i.e. it is not specifically and solely about the relevant conscious state [11]. The flip side of the argument is that introspection is not exhaustive, i.e. there may be more to the conscious state than what is captured by introspection. A fast response to both sides of this argument is that it confuses introspection with the report. The report may not be exhaustive and exclusive, as one, obviously, may not report all aspects or report too many. One may, however, speculate that the argument runs deeper. For instance, it might be the case that we do not introspect all of our experiences at the same time, but, as is the case with other functions of attention, only partial information is selected for further processing. If this were the case, one could ask participants to introspect on some aspects, but not all aspects, of an experience. This would in that case have great importance for how to study consciousness experimentally. The exact wording of instructions may have a great impact on the way a participant attends to her own experiences and which aspects are introspectively accessed.

As the individuation of different conscious contents and cognitive states and processes is an empirical matter, these speculations seem, at least in principle, to be open to empirical investigation and resolution as well. Conscious contents are individuated through the process of introspection, whereas cognitive states and processes are individuated by inference from behaviour. The considerations so far predict that different types of instructions and different types of report will give rise to different results in experiments on consciousness owing to a modulation of the 'kind of access' or 'target of introspective attention'. We will discuss these findings in the following.

## 2. EXPERIMENTS USING SUBJECTIVE REPORTS

Before discussing the most recent experimental findings, we provide a brief history of how such measures have been used in empirical science and how the quality of a measure might be evaluated.

### (a) *Subjective measures: methods and issues*

Subjective measures have been used in psychological science for more than 100 years. Possibly the first

study is reported by Sidis [12]. He presented participants with single letters or digits at various distances. The participants made (introspective) judgements as to whether or not they could see what was on the card followed by an attempt at identification. Sidis observed that even when participants claimed not to see the letter or digit, they performed above chance on the identification task. This kind of paradigm has also been referred to as the *subjective threshold* or *dissociation approach*, and has since been used in numerous studies, including very recent experiments [13]. Unconscious processing, in this case, is presumed to be responsible for any above-chance performance found when stimuli are below the so-called subjective criterion (i.e. when participants claim to have no experience of the identity of the stimulus) [14]. In a more recent variation of this, participants are asked instead to report their confidence in being correct. This method is referred to as establishing unconscious processing by 'the guessing criterion' [15].

Already, we see a difference between different scales of awareness. When using the scale introduced by Sidis, participants are asked only to introspectively report their experience, but not to draw any inferences about the accuracy of a classification process. On the other hand, when participants are asked to report their confidence in being correct (or in a slightly different paradigm described below, place a wager on being correct), they perform a metacognitive judgement ('how good was the classification?'), and they are presumed to use (only) their conscious experiences as the basis of this judgement [16]. One type of scale thus encourages introspection (judgement of the clarity of the experience) whereas the other type of scale encourage metacognition (judgement of the quality of the classification process). Various theoretical arguments have been made in favour of both types of scales [17], and in the following, we focus mainly on empirical differences. Both types of scales, however, use the dissociation approach as noted earlier.

There are certain problems with the dissociation approach. As a participant, when you see something so vaguely that you have almost no confidence at all in what you see, you may be reluctant to claim that you are in fact seeing something. If anything, this tendency would be expected to increase when you know that an experimenter will analyse your data to see if you are actually correct when you claim to see something. In other words, it should in fact be expected that participants in experiments are holding back information about very vague experiences or about classifications in which they have very little confidence. However, it is required of a subjective measure that it detects all relevant conscious knowledge (or all experiences) [11,18–20]. Technically, this has been referred to as *exhaustiveness* [11], and we might expect that different measures will differ in their degree of exhaustiveness.

Unfortunately, we cannot simply solve the problems associated with poor exhaustiveness by using the scale that shows the greatest sensitivity as some scales might misclassify unconscious processes as conscious and thus give 'opposite' results. If, for instance, we used a purely behavioural measure such as task accuracy to test for the presence of conscious experience—this is

indeed sometimes performed when experimenters want to make sure that no conscious experience is present [14,21]—we would risk classifying some unconscious information as conscious. If we are to trust our scale, it should not classify any unconscious information as conscious—that is, it should be *exclusive* [11]. As with exhaustiveness, we might expect different scales to differ in the extent to which they are exclusive.

Subjective measures of consciousness should obviously be as exhaustive and as exclusive as possible, yet it is not entirely clear how to go about this. As we cannot simply use the most sensitive scale (which might not be exclusive), we would have to compare only those scales that we have no reason to believe would mistake unconscious processing for conscious processing. If there is no *a priori* reason to suspect that a collection of scales are suboptimally exclusive, we can compare their sensitivity, and the scale that is the most sensitive will be the most exhaustive. This will thus be the preferred scale (all else being equal). Two common ways in which scales are compared are by (i) the amount of unconscious processing they indicate and (ii) how well and how consistently ratings correlate with accuracy? Under usual circumstances, for a scale to be maximally exhaustive, it should indicate as little subliminal perception as possible (i.e. participants are holding back the smallest amount of information), and the correlation between accuracy and awareness should be as good as possible (i.e. there is a large difference in accuracy when participants claim to see nothing and when they claim to have a clear experience). A further test for the trustworthiness of a measure is its stability, e.g. it is not that ‘seeing nothing’ is associated with one accuracy level one moment, and a completely different level the next. One or more of these methods have been used to compare different awareness scales in a number of experiments. These experiments are summarized below.

### (b) *The impact of scale steps*

Overall, the impact on exhaustiveness has been examined for two types of scale manipulations. First, modifications of the number and descriptions of scale steps have been examined, and second, changes in the type of scale (e.g. whether participants report consciousness, confidence or something else) has been examined. For the first manipulation, we are primarily aware of studies using purely introspective measures and thus not drawing upon other metacognitive skills (such as judgement of confidence in a classification process). For this reason, this section draws more heavily on purely introspective measures.

A number of experiments have found above-chance performance when participants claim to have no experience of the stimulus (i.e. when the stimulus is not perceived according to the subjective threshold as established by an introspective report). However, Ramsøy & Overgaard [2] drew attention to the fact that many such studies divide subjective experiences into ‘clear, vivid experiences’ and ‘nothing at all’, a division that might not capture the descriptions given by participants, and previous studies might thus have used inappropriate introspective measures. Participants often claim to experience stimuli in quite

different ways from trial to trial, and for that reason Ramsøy and Overgaard examined if any subliminal perception was found if participants used a scale that followed their own descriptions of experiences rather than a dichotomous all-or-none scale. Both in a pilot study and in the actual experiment, participants performed a visual discrimination task (forced-choice of position (three possible locations), form (three geometrical figures) and colour (three different colours) of stimulus) and subsequently reported their awareness. They were asked to construct their own awareness scale. Participants generally ended up using a 4-step scale with the following scale step labels ‘No experience’, ‘Brief glimpse’, ‘Almost clear experience’ and ‘Clear experience’ (this scale will be referred to as the perceptual awareness scale, or PAS, in future experiments). When participants used the ‘Brief glimpse’ category they reported no awareness of form, colour, or shape (but instead only a general vague experience of having seen something). A few participants included two additional steps, but these had no separate description and were rarely used.

The results demonstrated that each PAS rating was related to a different accuracy level, with accuracy increasing as a function of PAS rating. In other words, the individual subjective ratings corresponded to different levels of objective performance. Additionally, no statistically significant above-chance performance was found when participants used the scale step ‘no experience’. However, if the scale step for which participants claimed not to see stimulus features or location (i.e. the ‘Brief glimpse’ category) was also included, above-chance performance was indeed found as in previous studies. The study can be criticized for not comparing the results with the results obtained by an actual dichotomous scale and, partly for this reason, a second study was performed by Overgaard *et al.* [22]. Another purpose of the study was to examine if awareness was gradual in a general sense—i.e. if any feature can be perceived more or less clearly, or whether partial awareness is simply full awareness of individual features as has been hypothesized by others [23].

Data from Ramsøy & Overgaard [2] seemed to support the notion that awareness is gradual in a general sense—i.e. that any stimulus feature can be perceived in a gradual way. Even a line segment or a dot might be perceived more or less clearly. Yet, based on the data, it was difficult to rule out that the reports of partial awareness were caused by participants perceiving, for example, one line of a geometrical figure. For this reason, very simple stimuli were used in the 2006 study: participants were presented with small grey lines on a white background. Most grey lines were tilted 45° clockwise from vertical, but on each trial a small group of lines in one quadrant of the display were instead tilted 270°. This group of lines was the target, and the participants were asked to report in which quadrant the target appeared and subsequently rate their awareness of the target dichotomously or using the PAS.

The results replicated the earlier findings that accuracy increased as a function of PAS rating. Furthermore, accuracy was higher when participants rated a stimulus as unseen on the dichotomous scale

than when they rated it as unseen on PAS (35% versus 31%)—PAS thus proving more exhaustive than a dichotomous scale—and accuracy was found to be a lot lower when participants rated a stimulus as seen on the dichotomous scale than on PAS (78% versus 94%). If consciousness is always all-or-none, there would be little or no reason that these differences should be observed, and Overgaard *et al.* [22] thus concluded that there is evidence that consciousness is a gradual phenomenon even when very simple stimuli are used.

The difference between using a dichotomous measure of awareness and a 4-step measure (PAS) has also been examined in blindsight. Blindsight patients arguably report no visual experiences in a part of their visual field corresponding to a neural injury to V1, and thus consider themselves to be (partially) blind. Nevertheless, in certain laboratory tests, they perform well above chance when performing forced-choice tasks on visual stimulation in the blind field [24]. The common interpretation is that vision can occur in the complete absence of awareness, yet a few papers have reported the presence of weak visual experiences in blindsight [25,26]. Having observed that PAS seemed more exhaustive than a dichotomous measure, Overgaard *et al.* [27] examined whether different conclusions would be reached if a blindsight patient was tested with PAS rather than a dichotomous measure.

They examined a patient with damage to her left visual cortex and apparent corresponding blindness in her right visual field. In a first experiment, they presented her with letters in different parts of the visual field. She failed to report any letters displayed in the upper right quadrant in spite of successfully reporting seeing all letters presented elsewhere. In a second experiment, the patient was presented with geometrical figures (in the healthy as well as the injured field) and asked to guess which one was presented on each trial and subsequently rate her awareness on a dichotomous scale. The typical blindsight findings were replicated; on the trials on which the patient reported a stimulus as ‘not seen’ in her injured field she nevertheless performed well above chance (46% versus a chance level of 33%) and accuracy did not vary significantly as a function of awareness rating. However, when PAS was used in a third experiment, a strong relationship between awareness rating and accuracy was observed, and it seemed very similar to the relationship in the intact field. In other words, it was shown that at least for that particular blindsight patient, the observed above-chance performance was better explained by weak experiences than intact processing in the absence of awareness when PAS was used, thus indicating that the 4-point awareness scale with each step labelled by participants was more exhaustive than a dichotomous awareness scale.

We are aware of only a few studies examining the impact of different confidence scales on consciousness and these give somewhat mixed results. These studies all employ artificial grammar learning and test for conscious/unconscious knowledge. In this paradigm, participants are typically first asked to memorize a large number of letter strings. The participants are

subsequently told that each string obeyed one of the two complex rules, and in the second part of the experiment, they are presented with the strings again and for each string they are asked to indicate if they believe it obeyed rule a or b. After each classification, participants report their confidence, and the relationship between confidence and accuracy can be examined. Quite surprisingly, Tunney & Shanks [28] and Tunney [29] found that a binary ‘high/low’ scale was able to detect differences between conscious and unconscious processing, whereas a continuous scale from 50 to 100 (indicating the estimated accuracy from chance to complete certainty) was not. Both studies, however, used a very difficult task (accuracy approx. 55%). Dienes [30] hypothesized that the results might reflect that our judgements of confidence are not numerical as such, and converting our feelings of confidence to a number might be a noisy process.

Dienes [30] repeated the experiment using six different scales (a ‘high/low’ scale, a ‘guess/no guess’ scale, a ‘50–100’ scale with descriptions of what the numbers mean, a ‘50–100’ scale without descriptions of what the numbers mean, a ‘numerical categories 50–100’ scale where it is only possible to report 50, 51–59, 60–69, . . . , 90–99, 100 and finally a 6-step scale with verbal categories). Using the same stimuli as Tunney and Shanks, he found that overall, there was no difference between scales—the only scale seemingly (but not significantly) performing slightly worse was the numerical categories scale. He concluded that the type of scale made little difference, at least in a very difficult task. Using easier stimuli, a comparison between the sensitivity of a ‘high/low’ and a ‘50–100’ scale gave the opposite result of Tunney and Shanks, i.e. that a more fine-grained scale was more sensitive.

Taken together, the studies discussed above indicate that when introspecting, dichotomous scales are sub-optimal in many cases and the subjective measure in visual awareness tasks seems to benefit from allowing the participants to define the number of scale steps and their description (or at least using a scale that is created by participants in similar studies rather than one created arbitrarily by the experimenter). For scales using metacognitive judgements of classification performance, only artificial grammar learning has been examined. Here, it seems that when task accuracy is very low, using a fine-grained scale makes no difference (or possibly it makes the results worse), whereas a fine-grained scale appears useful when the task is somewhat easier. In the following, we will discuss the research into scale comparison, i.e. whether a scale using purely introspection performs better or worse than the scales using metacognitive judgements about task accuracy.

### (c) *The impact of scale type*

In different experiments, different measures of awareness have been used, yet very few studies have examined which of the scales in use is the most exhaustive. When examining conscious experience, the most intuitive thing to ask about is probably just that, conscious experience (i.e. asking participants to introspect on their experience). At least, this is the oldest

method, and it is still used very frequently today (in the PAS experiments, for example). However, as mentioned earlier, introspection has often been criticized as inaccurate, and many scientists prefer measures that are functionally defined. Confidence ratings (CRs) have been used as an alternative to ratings of experience in part because they do not ask participants to rate their experience as such, but instead ask participants about their insight into a cognitive process. Additionally, such measures can be used across many different paradigms (e.g. the same confidence scale can be used whether the participant is viewing geometrical figures, viewing motion or even performing artificial grammar learning). CRs have been used either with respect to perception itself, in which case participants report their confidence in having perceived something (note that this type of scale has some introspective qualities) [31,32], or with respect to participants' performance, in which case they report their confidence in having provided a correct answer [15,30,33].

Asking participants to place a wager on their classification has been used as an alternative to ratings of experience or confidence. Originally, gambling was used in cases where the participants could not be expected to understand a confidence scale; Ruffman *et al.* [34] used it with 3-year-olds, and Shields *et al.* [35], Kornell *et al.* [36] and Kiani & Shadlen [37] used it with rhesus monkeys. Recently, Persaud *et al.* [16] used post-decision wagering (PDW) with adult humans (normal participants and a blindsight patient), and they argued that it should be the preferred method because it required, according to them, no introspection (i.e. it was claimed to be an objective measure) and the prospect of making money motivates participants to reveal all knowledge. PDW was thus claimed to be the measure with the highest degree of exhaustiveness, yet no direct comparison was made regarding exhaustiveness in terms of performance at the subjective threshold and the correlation between accuracy and wagers/awareness rating. The claim that PDW is an objective measure was quickly challenged by Seth [38], who argued that PDW required metacognitive judgements just like any confidence scale, and Clifford *et al.* [39] argued that the unconscious processing indicated in the experiments of Persaud *et al.* was likely to be a consequence of whatever criterion the participants set for when to wager high (in other words, the results could be caused by suboptimal exhaustiveness). Additionally, Sandberg *et al.* [17] drew attention to the fact that the use of PDW with real money alters the performance of the objective classification task.

In order to empirically test the claims made in the PDW papers, Dienes & Seth [40] compared a PDW scale with another metacognitive measure that did not encourage participants to introspect directly (a confidence scale) in the context of an artificial grammar-learning paradigm. Dienes and Seth wanted to examine whether PDW is indeed a better or more objective measure of awareness than CR scales when participants belong to a population that is expected to be able to understand and use CR scales (in this case, adult humans with well-developed linguistic abilities). In addition to simply comparing

the scales, they administered a test of risk aversion in order to examine whether PDW was more closely linked to risk aversion than CR.

Dienes and Seth performed two experiments. In their first experiment, they simply asked participants to assign letter strings to one of the two categories and either rate their confidence in being correct or wager one or two sweets. Although CR performed numerically better, they found no statistically significant difference between groups for the amount of unconscious processing at the subjective threshold or for the correlation between CR and accuracy. They also examined a different measure closely related to confidence–accuracy correlations, type 2d', and found no difference here either. Interestingly, however, they did observe a negative correlation between risk aversion and type 2d' and marginally so between risk aversion and confidence–accuracy correlation when participants used wagering, but not when they used CRs. In other words, the more risk averse you are, the smaller the estimated amount of conscious knowledge as measured by a wagering scale, but not by a confidence scale.

In their second experiment, Dienes and Seth changed their wagering procedure so participants could no longer lose anything. After making their classification, participants could choose to stick to their decision and gain a sweet if they were correct or simply randomly determine if they would get a sweet or not by drawing a card (50% probability). In this case, the lower step on the scale is clearly associated with complete chance performance and, in this way, it can be said that participants are instructed to use this scale step only when they believe they are completely at chance—i.e. the scale now resembles a standard confidence scale. Not surprisingly, this manipulation resulted in participants using the scale very similarly to how the confidence scale was used in experiment 1, and the correlation between risk aversion and conscious knowledge disappeared. The correlation between accuracy and awareness was also marginally higher than for PDW in experiment 1 (but no different from for CR). The overall conclusion of Dienes and Seth's study is thus that PDW does not show superior performance in artificial grammar-learning paradigms, and if scientists want to avoid risk aversion influencing the experimental outcome, then they may have to alter the instructions to make the PDW scale very similar to a confidence scale. For this reason, a confidence scale seems preferable compared with a PDW scale whenever participants can be expected to be able to use such a scale.

In contrast to Dienes and Seth, Sandberg *et al.* [17] used a visual paradigm and they not only examined confidence about being correct and wagering, but also report the perceptual experience as revealed introspectively on the PAS. Participants were asked to discriminate briefly presented geometrical figures (four choices), and subsequently rate their awareness on one of the three scales (PDW, CR and PAS). All scales had four scale steps. Interestingly, Sandberg *et al.* [17] found that PAS indicated the lowest accuracy at the lowest scale step. An accuracy level of 27.9 per cent was found, only just significantly above chance, 25 per cent, when uncorrected for multiple

comparisons. In comparison an accuracy of 36.6 per cent was found for CR, and 42 per cent was found for PDW. The correlation between accuracy and awareness rating was also examined for all scales. Again, PAS gave the best results; the best overall correlation was found, and the ratings were used more consistently across stimulus durations (i.e. each awareness rating was more consistently related to a particular accuracy rating than in other scales). The experiment thus confirmed the findings of Dienes & Seth [40] that CR performs similar to or slightly better than PDW, but in a different paradigm. Additionally, the experiment showed that introspective reports of awareness performed better than either of the other two scales.

One possible reason for the results could be the response distribution. When participants use PAS, they use all scale steps roughly the same number of times, whereas participants used PDW and to some extent CR more dichotomously. This results in awareness ratings of 1 and 4 being used across many different accuracy levels, thus giving a worse correlation and poor exhaustiveness at awareness ratings of 1. The failure to increase wagers can be explained by risk aversion as shown by Dienes & Seth [40], yet this cannot explain why PAS performs better than CR. One straightforward, yet somewhat controversial (cf. Dienes & Seth [41] and Timmermans *et al.* [42]) explanation is that participants are quite able to report small differences in experience, but they have no knowledge that these small differences are significant enough to alter performance. In this context, it is interesting to note that CR and PDW require metacognitive insight into the classification process, which is exactly what Nisbett and Wilson showed to be unreliable, whereas PAS does not. In other words, reporting confidence in the response might be considered a harder task than reporting awareness of the stimulus, and performing this task optimally requires some degree of introspection along with a successful additional metacognitive process of relating experience and accuracy, a skill that participants might not always possess. This hypothesis is not easily tested, but experimental designs might be proposed drawing on the fact that metacognition is a skill (individual differences and specific neural correlates are found [1,43]). Thus, if CR and PDW reports tax metacognitive skills to a higher degree than introspection, then they might be affected more by distracter tasks or pressure to report quickly.

An introspective measure has also been compared with PDW by Nieuwenhuis & de Kleijn [44]. They examined the attentional blink in four experiments, using an awareness scale and wagering. During the attentional blink, the ability to detect a target stimulus, T2, is impaired by presenting it shortly after another target, T1, in a series of rapidly presented stimuli. Like Overgaard *et al.* [22], Nieuwenhuis and de Kleijn wanted to examine the claim that consciousness is an all-or-none phenomenon. Sergent & Dehaene [23] had found that this seemed to be the case, at least, during the attentional blink, and so far, a continuous transition from unconscious to conscious processing had not been found in this paradigm.

Sergent and Dehaene used a 21-point scale, which has been criticized for being confusing for participants [22]. Nieuwenhuis and de Kleijn reduced the number of scale steps to 7 (still an arbitrary, although lower number) in their first experiment, and replicated the findings of Sergent and Dehaene. When they used wagering in their second experiment, however, the dichotomous response pattern disappeared to some extent. This is somewhat surprising as wagering was found to be a poor measure in the above-mentioned experiments. Inspection of the tasks as well as the perceptual awareness scale and the wagering scale reveals some explanation for this.

The attentional blink is, in most cases, presented as a detection task, not as an identification task, which is the type of task for which the PAS, for instance, is developed. Nieuwenhuis and de Kleijn asked participants to perform a discrimination task for T1 while simply reporting the clarity of T2, which may or may not be present (when not present, a blank slide was shown instead). In this case, the awareness rating and the wagering response relate to a presence/absence judgement (a detection task), which is nevertheless not explicitly performed. For wagering, participants were allowed to wager three different amounts on the absence or presence of T2 and even not to wager, whereas awareness ratings were made on a single scale from 'not seen' to 'maximal visibility'. The lowest awareness rating 'not seen' thus covered anything from complete certainty that nothing was displayed (e.g. completely clear perception of the empty slide) to no awareness of anything. In other words, the lowest step on the perceptual awareness scale corresponded to a combination of four steps on the wagering scale. The two scales, in this sense, were not comparable, and it seems there is a need to construct a perceptual awareness scale that can be used in detection tasks. All this considered, Nieuwenhuis and de Kleijn, nevertheless, demonstrated a continuous transition between conscious and unconscious processing in their second experiment. However, as no attentional blink was found in their second experiment, they performed two additional experiments for which task difficulty was increased by changing both target single-digit numbers and distractors to single letters. In these experiments, participants were also asked to identify both T1 and T2 after rating their awareness or placing their wager (still based on present/absent judgements). With increased task difficulty, a continuous transition pattern was found for both wagering and perceptual awareness scales. In this case, a PDW scale seemed to perform as well as or slightly better than an awareness scale with an arbitrary number of steps, which was not generated with a detection task in mind.

#### (d) *The impact of stimulus intensity*

Interestingly, all types of scales (introspective, confidence or wagering) indicate that unconscious processing occurs at very specific stimulus intensities. In the experiment by Sandberg *et al.* [17], they found that stimuli had to be presented for 50 ms for participants to be able to identify them at a rate above chance. When stimuli were presented for around 130 ms (or more), participants were

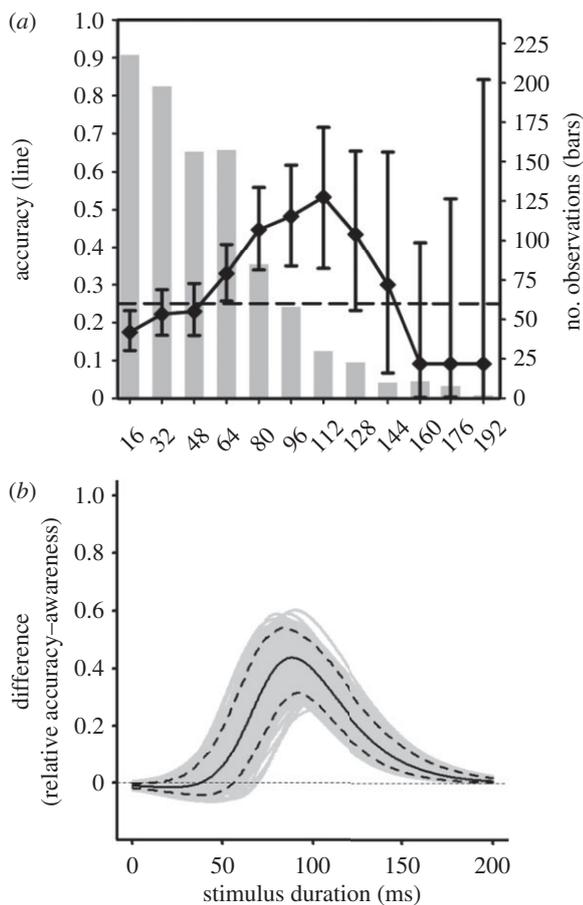


Figure 1. A window of subliminal perception. Estimated subliminal perception across stimulus duration in a visual identification task across 12 participants using perceptual awareness scale using two different methods of analysis. Unconscious perception can be calculated using the subjective threshold. Here, unconscious perception accounts for any above-chance performance when the participant claims to have no experience of the stimulus. Unconscious perception can also be estimated from relative differences between average accuracy and awareness responses [45]. This method built on the fact that task accuracy and awareness can be described as sigmoid functions of stimulus duration, i.e. what has been called a psychometric and a conscious-metric curve [46]. Using this method, unconscious perception is calculated by subtracting sigmoid accuracy functions from sigmoid awareness functions. (a) Subliminal perception calculated using the subjective threshold. Error bars indicate 95% confidence intervals (CI) calculated from binomial distributions. (b) Subliminal perception calculated by subtracting the sigmoid awareness function from the sigmoid accuracy function. Note that both methods of analysis indicate unconscious perception to occur only for stimulus durations of around 50–150 ms.

able to identify them with an accuracy of almost 100 per cent. Unconscious perception, as indicated by subjective threshold analysis (task accuracy when participants claim to have no awareness), is plotted in figure 1*a*. Here, it can be seen that all unconscious processing is found within this time window, i.e. unconscious perception starts when performance deviates from chance and disappears again shortly after peak performance is reached. However, one large problem for the analysis at the subjective threshold is that it is difficult to conclude much about unconscious processing at high stimulus

intensities. The reason for this is that only awareness ratings of 1 are used in the analysis, and when stimulus intensity is very high, participants rarely claim not to see anything (i.e. the number of observations dropped drastically when stimuli were presented for more than 100 ms as indicated by the bars in figure 1*a*). Yet, this possible ‘window of subliminal perception’ was confirmed by analysing the data in a different way.

Sandberg *et al.* [45] drew attention to the fact that accuracy and awareness in visual discrimination tasks can be described as sigmoid functions of stimulus intensity as had first been hypothesized by Koch & Preusschoff [46]. It has long been known that average task accuracy can be described as a sigmoid function (a standard psychometric curve), but Sandberg *et al.* [45] found this to be the case for awareness ratings as well. By fitting sigmoid functions to the average accuracy and awareness (taking into account variability between participants), group estimates of accuracy and awareness functions could be made. Sandberg *et al.* [45] found that in general, the awareness function lagged the accuracy function, which could be taken as an indication of unconscious perception. Their analysis confirmed that awareness was generally lagging accuracy, and that awareness increased more slowly than accuracy. In other words, accuracy and awareness functions start increasing from the bottom plateau at roughly the same time, yet the awareness function increases much more slowly.

Because every data point (i.e. accuracy and awareness rating for all trials, not just the ones where participants claimed no awareness) was used in the analysis, the overall CI with which the sigmoids were estimated were very small compared with the CI of subjective threshold approaches, and this allowed for additional analyses. In order to estimate the stimulus durations for which the awareness was lagging accuracy most clearly, the awareness function was subtracted from the accuracy function. The result of this subtraction is shown in figure 1, and it is clear that this method reveals a ‘window of subliminal perception’ that is quite similar to that found using the subjective threshold approach (although the CI are somewhat narrower for the curve estimations and the curve is smoother).

### 3. CONCLUDING DISCUSSION

Summing up the above context, it appears that, at least for visual discrimination tasks, the best results are obtained using the measure drawing most heavily on introspection, i.e. the PAS, but not other metacognitive skills. In the one experiment [17] comparing PAS with CRs and wagering, PAS indicated less subliminal perception as well as a better and more stable correlation between accuracy and awareness, whereas wagering performed the worst. This seems to indicate that PAS is unaffected by risk aversion (as are CRs), and possibly the introspective task of reporting perceptual clarity is easier for participants than estimating accuracy, which may require both perceptual clarity as well as metacognitive knowledge of how this corresponds to different accuracy levels. At present, it is unclear whether introspection-based scales generated by participants

also perform better in artificial grammar-learning and attentional blink experiments.

Non-dichotomous introspective scales give more exhaustive results than dichotomous. They indicate less subliminal perception and they indicate a better correlation between accuracy and awareness. The scale steps on the 4-point PAS each correspond to accuracy levels that remain fairly stable across different conditions within the same experiment. Small amounts of unconscious processing seem to be found no matter which scale is used as the measure of awareness (although, in the early PAS experiments as well as in the blindsight experiment, no significant above-chance performance was found).

In the review of recent findings using direct, introspective reports, it thus seems evident that the way in which participants are instructed to report has significant impact on results. This, we believe, is empirical support of the claim that introspective reports should be considered as distinct from other kinds of metacognitive reports, and that the distinctions shown in table 1 are upheld. Further studies are necessary to conclude how strongly these distinctions should be interpreted.

In one interpretation, the distinction is represented at a neural level and captures a natural distinction of mental states. This version is cautiously suggested in Overgaard *et al.* [22], who report different neural activation patterns when asking participants to report introspectively about visual experiences in contrast to asking them to report in a non-introspective way (i.e. without attending directly to their experiences). However, a different interpretation would refrain from ontological commitments and limit the distinction to the methodological domain. According to this weaker version, we get different results with different instructions not because of distinctions in nature, but solely because of the differences in methodology. Although this conclusion immediately seems simpler to draw, it does demand a different ontology able to account for the empirical findings.

Regardless of the choice of interpretation, we believe that we have shown that some distinctions must be drawn between cognitive and conscious states, metacognition and introspection and that these distinctions have important implications for how to think about and experiment on the mind.

M. O. and K.S. were supported by the European Research Council.

## REFERENCES

- Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J. & Rees, G. 2010 Relating introspective accuracy to individual differences in brain structure. *Science* **329**, 1541–1543. (doi:10.1126/science.1191883)
- Ramsøy, T. Z. & Overgaard, M. 2004 Introspection and subliminal perception. *Phenomenol. Cogn. Sci.* **3**, 1–23. (doi:10.1023/B:PHEN.0000041900.30172.e8)
- Rounis, E., Maniscalco, B., Rothwell, J., Passingham, R. & Lau, H. 2010 Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci.* **1**, 165–175. (doi:10.1080/17588921003632529)
- Neisser, U. 1967 *Cognitive psychology*. New York, NY: Meredith.
- Shimamura, A. P. 2000 Toward a cognitive neuroscience of metacognition. *Conscious. Cogn.* **9**, 313–323; discussion 324–326. (doi:10.1006/ccog.2000.0450)
- Lyons, W. 1986 *The disappearance of introspection*. Cambridge, MA: MIT Press.
- Marcel, A. 2003 Introspective report: trust, self knowledge and science. *J. Conscious. Stud.* **10**, 167–186.
- Overgaard, M. 2006 Introspection in science. *Conscious. Cogn.* **15**, 629–633. (doi:10.1016/j.concog.2006.10.004)
- Nisbett, R. E. & 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)
- Wilson, T. D. 2003 Knowing when to ask: introspection and the adaptive unconscious. *J. Conscious. Stud.* **10**, 131–140.
- Reingold, E. M. & Merikle, P. M. 1988 Using direct and indirect measures to study perception without awareness. *Percept. Psychophys.* **44**, 563–575. (doi:10.3758/BF03207490)
- Sidis, B. 1898 *The psychology of suggestion*. New York, NY: Appleton.
- Schwiedrzik, C. M., Singer, W. & Melloni, L. 2011 Subjective and objective learning effects dissociate in space and in time. *Proc. Natl Acad. Sci. USA* **108**, 4506–4511. (doi:10.1073/pnas.1009147108)
- Snodgrass, M. & Shevrin, H. 2006 Unconscious inhibition and facilitation at the objective detection threshold: replicable and qualitatively different unconscious perceptual effects. *Cognition* **101**, 43–79. (doi:10.1016/j.cognition.2005.06.006)
- Dienes, Z., Altmann, G. T. M., Kwan, L. & Goode, A. 1995 Unconscious knowledge of artificial grammars is applied strategically. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 1322–1338. (doi:10.1037/0278-7393.21.5.1322)
- Persaud, N., McLeod, P. & Cowey, A. 2007 Post-decision wagering objectively measures awareness. *Nat. Neurosci.* **10**, 257–261. (doi:10.1038/nn1840)
- Sandberg, K., Timmermans, B., Overgaard, M. & Cleeremans, A. 2010 Measuring consciousness: is one measure better than the other? *Conscious. Cogn.* **19**, 1069–1078. (doi:10.1016/j.concog.2009.12.013)
- Merikle, P. M. 1982 Unconscious perception revisited. *Percept. Psychophys.* **31**, 298–301. (doi:10.3758/BF03202538)
- Merikle, P. M. & Joordens, S. 1997 Measuring unconscious influences. In *Scientific approaches to consciousness* (eds J. Cohen & J. Schooler), pp. 109–123. Mahwah, NJ: Erlbaum.
- Reingold, E. M. & Merikle, P. M. 1990 On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind Lang.* **5**, 9–28. (doi:10.1111/j.1468-0017.1990.tb00150.x)
- Snodgrass, M. 2002 Disambiguating conscious and unconscious influences: do exclusion paradigms demonstrate unconscious perception? *Am. J. Psychol.* **115**, 545–579. (doi:10.2307/1423527)
- Overgaard, M., Rote, J., Mouridsen, K. & Ramsøy, T. Z. 2006 Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Conscious. Cogn.* **15**, 700–708. (doi:10.1016/j.concog.2006.04.002)
- Sergent, C. & Dehaene, S. 2004 Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.* **15**, 720–728. (doi:10.1111/j.0956-7976.2004.00748.x)
- Sanders, M. D., Warrington, E. K., Marshall, J. & Weiskrantz, L. 1974 'Blindsight': vision in a field defect. *Lancet* **1**, 707–708. (doi:10.1016/S0140-6736(74)92907-9)

- 25 Stoerig, P. & Barth, E. 2001 Low-level phenomenal vision despite unilateral destruction of primary visual cortex. *Conscious. Cogn.* **10**, 574–587. (doi:10.1006/ccog.2001.0526)
- 26 Zeki, S. & Ffytche, D. H. 1998 The Riddoch syndrome: insights into the neurobiology of conscious vision. *Brain* **121**, 25–45. (doi:10.1093/brain/121.1.25)
- 27 Overgaard, M., FehI, K., Mouridsen, K., Bergholt, B. & Cleeremans, A. 2008 Seeing without seeing? Degraded conscious vision in a blindsight patient. *PLoS ONE* **3**, e3028. (doi:10.1371/journal.pone.0003028)
- 28 Tunney, R. J. & Shanks, D. R. 2003 Does opposition logic provide evidence for conscious and unconscious processes in artificial grammar learning? *Conscious. Cogn.* **12**, 201–218. (doi:10.1016/S1053-8100(02)00068-5)
- 29 Tunney, R. J. 2005 Sources of confidence judgments in implicit cognition. *Psychon. Bull. Rev.* **12**, 367–373. (doi:10.3758/BF03196386)
- 30 Dienes, Z. 2008 Subjective measures of unconscious knowledge. *Prog. Brain Res.* **168**, 49–64. (doi:10.1016/S0079-6123(07)68005-4)
- 31 Bernstein, I. H. & Eriksen, C. W. 1965 Effects of ‘subliminal’ prompting on paired-associate learning. *J. Exp. Res. Pers.* **1**, 33–38.
- 32 Cheesman, J. & Merikle, P. M. 1986 Distinguishing conscious from unconscious perceptual processes. *Can. J. Psychol.* **40**, 343–367. (doi:10.1037/h0080103)
- 33 Scott, R. B. & Dienes, Z. 2008 The conscious, the unconscious, and familiarity. *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 1264–1288. (doi:10.1037/a0012943)
- 34 Ruffman, T., Garnham, W., Import, A. & Connolly, D. 2001 Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *J. Exp. Child Psychol.* **80**, 201–224. (doi:10.1006/jecp.2001.2633)
- 35 Shields, W. E., Smith, J. D., Guttmanova, K. & Washburn, D. A. 2005 Confidence judgments by humans and rhesus monkeys. *J. Gen. Psychol.* **132**, 165–186.
- 36 Kornell, N., Son, L. K. & Terrace, H. S. 2007 Transfer of metacognitive skills and hint seeking in monkeys. *Psychol. Sci.* **18**, 64–71. (doi:10.1111/j.1467-9280.2007.01850.x)
- 37 Kiani, R. & Shadlen, M. N. 2009 Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* **324**, 759–764. (doi:10.1126/science.1169405)
- 38 Seth, A. K. 2008 Post-decision wagering measures metacognitive content, not sensory consciousness. *Conscious. Cogn.* **17**, 981–983. (doi:10.1016/j.concog.2007.05.008)
- 39 Clifford, C., Arabzadeh, E. & Harris, J. A. 2008 Getting technical about awareness. *Trends. Cogn. Sci.* **12**, 54–58. (doi:10.1016/j.tics.2007.11.009)
- 40 Dienes, Z. & Seth, A. 2009 Gambling on the unconscious: a comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Conscious. Cogn.* **19**, 674–681. (doi:10.1016/j.concog.2009.09.009)
- 41 Dienes, Z. & Seth, A. K. 2010 Measuring any conscious content versus measuring the relevant conscious content: comment on Sandberg *et al.* *Conscious. Cogn.* **19**, 1079–1080; discussion 1081–1083. (doi:10.1016/j.concog.2010.03.009)
- 42 Timmermans, B., Sandberg, K., Cleeremans, A. & Overgaard, M. 2010 Partial awareness distinguishes between measuring conscious perception and conscious content: reply to Dienes and Seth. *Conscious. Cogn.* **19**, 1081–1083. (doi:10.1016/j.concog.2010.05.006)
- 43 Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S. & Rees, G. 2011 Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Conscious. Cogn.* **20**, 1787–1792. (doi:10.1016/j.concog.2010.12.011)
- 44 Nieuwenhuis, S. & Kleijn, R. 2010 Consciousness of targets during the attentional blink: a gradual or all-or-none dimension? *Atten. Percept. Psychophys.* **73**, 364–373. (doi:10.3758/s13414-010-0026-1)
- 45 Sandberg, K., Bibby, B. M., Timmermans, B., Cleeremans, A. & Overgaard, M. 2011 Measuring consciousness: task accuracy and awareness as sigmoid functions of stimulus duration. *Conscious Cogn.* **20**, 1659–1675. (doi:10.1016/j.concog.2011.09.002)
- 46 Koch, C. & Preusschoff, K. 2007 Betting the house on consciousness. *Nat. Neurosci.* **10**, 140–141. (doi:10.1038/nn0207-140)