Forthcoming in U. Kriegel (ed.), Handbook of the Philosophy of Consciousness, OUP

Consciousness, introspection, and subjective measures

Maja Spener
(University of Birmingham)

1. Introduction

In recent years, the divide between conscious and unconscious perception has been the focus of a thriving research area in the science of consciousness. Related topics include the existence of unconscious, 'subliminal' perception, graded conscious perception, and the threshold between conscious and unconscious perceptual processing. Subjective measures of consciousness play a prominent role in this literature and they are frequently referred to as 'introspective measures'. Subjective measures of consciousness have been the subject of intense criticism. Indeed, as some critics insist, they are in effect rehearsing century-old objections to the use of introspection in the empirical study of consciousness (Irvine 2012: 629).

This chapter discusses the main types of so-called 'subjective measures of consciousness' used in current-day science of consciousness. I explain the key worry about such measures, namely the problem of a putatively ever-present response bias. I then turn to the question of whether subjective measures of consciousness are introspective. I show that there is no clear answer to this question, as proponents of subjective measures do not employ a worked-out notion of subjective access. In turn, as I explain, this makes the problem of response bias less tractable than it might otherwise be.

2. A methodological challenge in the science of consciousness

The science of consciousness faces the following challenge: getting primary - i.e. most immediate or direct - data about consciousness seems to require doing so by relying on individuals' first-person access to their own experience[1]; but scientifically rigorous investigation requires embedding such data-acquisition in the empirical setting of experimental psychology (see, e.g. Overgaard 2006: 629). Currently, there are no so-called *objective measures of consciousness* - no thoroughly third-person observational measures - which are widely agreed to capture conscious aspects of the mind, rather than some behavioural capacity that may be correlated with phenomenal

[1] The term 'experience' and its cognates are used here to refer to mental episodes which are conscious.

consciousness. More particularly, there are no such measures which provide access to properties of consciousness *qua* conscious properties, or, *qua* conscious contents. We do not have, as some put it, a 'consciousness meter' (Timmermans and Cleeremans 2016: 21) which scientists can use to directly procure the needed data about consciousness. Plenty of empirical work employs objective measures of consciousness, e.g. on different levels of consciousness and on neural correlates of consciousness. Yet, such work invariably depends on assuming bridging hypotheses about the putative objective markers of consciousness (e.g. neural activity held to be associated with consciousness) that provide the link between them and conscious contents (Chalmers 2004, Irvine 2013, Phillips 2015a). Objective measures of consciousness are therefore not only indirect, they are also based on substantial assumptions concerning consciousness itself: certain behavioural or neural properties are assumed to be good markers of conscious content.

As generally acknowledged, people seemingly have immediate access to their own conscious contents from the first-person vantage point, i.e. from the subject's perspective. Some scientists therefore believe that the best way to measure conscious content must utilise subjects' apparently direct epistemic access to their experiences (Timmermans and Cleeremans 2016: 3). *Subjective measures of consciousness* thus involve the experimenter eliciting data about consciousness by getting subjects to provide it via reports under experimental conditions. These reports - often referred to as 'subjective reports' - are verbal or non-verbal responses given by subjects who are under explicit instructions to give certain kinds of information in relation to a perceptual task. Subjective reports are taken to be directly about, or significantly reflective of, the subject's experiential situation from their own, first-person perspective.

In relying on subjective reports to provide access to consciousness, subjective measures do not seem open to independent validation in the form of inter-measurement calibration (cf. Spener 2015). The lack of inter-measurement checks is held to be problematic because there are serious further worries about the reliability of subjective measures, namely that they are reflective of the subject's own views about consciousness, rather than consciousness itself. As a consequence, such measures have faced sharp criticism for apparent failure to comply with sufficiently rigorous scientific standards.

The appearance that subjective measures of consciousness are indispensable and yet scientifically inadequate, minimally presents a difficult challenge for the science of consciousness. There is debate over whether this challenge constitutes a fundamental and intractable problem for the field. Upbeat voices argue that the problems are significant but not fundamentally unsolvable. Overgaard (2016) e.g., considers the challenge a sign of the infancy of the science of consciousness insofar as the latter lacks standardization of its methods. But he is optimistic about finding such standards, in the same way other scientific fields have eventually succeeded in doing so. By contrast, Irvine (2012) holds that subjective measures in the science of consciousness are problematic in a more substantial manner and that these problems are distinctive to the field. She recommends getting rid of subjective measures altogether (Irvine 2012: 630-1, 642). Other sceptical views are more mixed in their message. Schwitzgebel (2011), e.g., offers those working in the science of consciousness a choice: either to operationalise consciousness in terms of some behavioural capacity, or to 'muddle on' with subjective measures (see Spener 2013, Schwitzgebel 2013 for further discussion).

3. Subjective and objective thresholds of perception

Developments in the science of consciousness - especially in the debate about subliminal perception - over the last sixty years are a testament to the persisting challenge. Throughout, scientists show recognition of serious problems with subjective measures of consciousness, but the overall trends also indicate a lack of genuinely viable alternatives to them. The problematic situation is manifested by the lack of agreement about how to define and where to locate conscious perception in the first place. Eriksen (1956, 1960)'s influential critique of subjective measures of consciousness is a milestone in the debate. Eriksen argued that subjective reports - the key ingredient in subjective measures - are reflective of subjects' response bias and that this undermines their suitability to measure the genuine onset of conscious perception.

Eriksen drew on (then relatively new) work involving the application of signal detection theory (SDT) to cases of perception-based human decision-making (Green and Swets 1966). SDT is a statistical framework which allows one to separately estimate both, the sensitivity of a response-making system to detect a stimulus in the presence of noise (its *objective sensitivity* to the stimulus), and the criterion the system uses for deciding when to respond that a signal is present rather than noise (its *decision criterion*). SDT assumes that in the case of signals under uncertainty, any response by the system is always a product of perceptual sensitivity and decision criterion.[2]

Perceptual detection invariably involves, and is thus modulated by, a noisy context (the more noise compared to the stimulus signal strength, the less sensitive a detection system is). As soon as noise is present, a system's detection response to a signal involves a decision in the face of some uncertainty about whether the stimulus is present. Such decision-making requires the system to set criteria, where these are, roughly, benchmarks for how strong a signal must be to merit a certain detection response.

These criteria can introduce considerable bias into decision-making, in that they may be asymmetric (e.g. manifest a preference for one response option over others) or not stable (e.g. allow for shifting thresholds for specific response options). The notion of *response bias* 'measures the participant's tilt towards one response or another' (Macmillan and Creelman 1991: 24).[3] For instance, two subjects may differ in their decision criteria under the same stimulus conditions by being liberal and conservative, respectively, in setting their threshold for a positive response to a signal. The first subject might use a criterion that aims to maximize correct positive responses and not be worried about getting a high number of incorrect positive ones in the process. The second subject might be guided in part by minimizing incorrect positive responses. Moreover, a single subject can be easily manipulated to change their own criterion (Green and Swets 1966). Raising the stakes for the subject to get the decision right in various ways - by imposing costs for positive responses when the signal is not present (punishing false alarms) - means that the subject will become more conservative in their decision-making, or less trigger-happy, so to speak. By lowering the stakes and increasing benefits of right answers (rewarding hits), the subject will become more liberal in responding, trying to maximise rewards.

To get an estimate of perceptual sensitivity *per se*, one needs to set aside response bias in the

[2] For classic and introductory texts on STD applied to human decision-making, see (Green and Swets 1966, MacMillan and Creelman 1991).

[3] For a discussion of structural biases inherent in different kinds of free choice tasks, e.g. yes/no or same/difference tasks see Phillips (2015b).

response data first. SDT offers a mathematical framework for how to do this, called a 'receiver operating characteristic (ROC) curve analysis', which involves taking account of the relationship between the rate of accurate positive responses (hit rate) and the rate of false positive responses (false alarm rate).[4]

The key lesson from SDT applied to human perception-based decision-making is that one cannot automatically read people's objective perceptual sensitivity to a stimulus off the rate of accurate responses. One first needs to filter out any bias in the responses (via ROC curve analysis). SDT is a model of perceptual sensitivity and decision-making, and it does not itself speak to how consciousness figures in the perceptual process. Eriksen's criticism is that when subjective measures are used to detect the presence or absence of conscious perception, they provide data directly derived from subjects' reports and such data is shown by SDT to be contaminated by response bias. He argued, moreover, that at-chance detection performance indicates that the stimulus at that level of intensity is not perceived at all, and that it is above-chance detection performance (measured objectively) that indicates conscious perception of the stimulus.

The first claim - that total lack of ability to detect a stimulus constitutes good evidence that it is not perceptually processed - is widely accepted. The second claim about the onset of consciousness is contentious. Cheesman and Merikle (1984, 1986) reject it, arguing that in addition to the threshold marked by at-chance detection performance (the *objective threshold*), there is an important report-based threshold (the *subjective threshold*), marked by the point of stimulus intensity at which subjects claim not to see the stimulus or to be guessing in relevant tasks. According to them, the this 'threshold of claimed awareness' marks the onset of conscious perception because it 'better captures the phenomenological distinction between conscious and unconscious perceptual experiences' (Merikle and Cheesman 1986: 344). On their view, since it can be shown that the subjective threshold is higher along the stimulus intensity dimension than the objective threshold, unconscious perception is perceptual processing *in between* the two thresholds, i.e. above the threshold for perceptual processing and below the threshold for conscious perceptual awareness.

It is easy to see that the objective threshold is implicit in the SDT model, mapping onto the estimate for objective perceptual sensitivity. As the model makes clear, though, whatever else subjective thresholds are indicative of, they are also reflective of subjects' response bias. Cheesman and Merikle (1984) note that when one uses subjects' reports about whether they are able to detect a stimulus to measure the boundary between conscious and unconscious perception, one in effect 'allows subjects to set their own thresholds'. Under the same stimulus conditions, two individuals may differ in their subjective thresholds merely because one is cautious and employs conservative decision criterion, resulting in a higher threshold of claimed awareness, and the other is rash and employs a liberal decision criterion, resulting in a lower threshold. To accommodate this worry, Cheesman and Merikle proposed that results concerning the distinction between conscious and unconscious perception that are based on subjective measures, must be confirmed by an additional criterion. Specifically, they propose that there must be evidence that the postulated different perceptual processes have a qualitatively different behavioural effect (Cheesman and Merikle 1986: 363).

[4] The details do not matter here. See Macmillan and Creelman 1991:45-70.

4. Response bias and subjective report

Response bias is often held to be a methodological affliction of specifically subjective measures of consciousness. This is because they are report-based measures and reports are problematic (see also, e.g. Irvine 2012: 641, Tunney and Shanks 2003: 1061):

> "...contrary to many researchers' implicit assumptions, (SDT shows) there is no such thing as an unmediated 'subjective report' - ever." (Snodgrass and Lepisto 2007: 526)

> "The signal detection theory perspective dictates that there is no such thing as a raw report uncontaminated by decision processes." (Block 2008: 312)

What exactly is the concern about reports and subjective measures relying on them, which is brought to light by SDT? To assess this, bear in mind two points. Firstly, response bias concerns responses in the context of perceptual detection tasks. When criticising Marcel (1980, 1983)'s results about subliminal perception from masking experiments, for instance, Holender (1986) and Cheesman and Merikle (1984, 1986), pointed out that yes/no tasks have been shown to induce systematic response bias by the subject. Another frequently used task, involving same/difference judgements, is skewed towards the 'same' option in conditions where discrimination performance gets nearer to chance.

Secondly, response bias is not pervasive. A biased response is due to the subject's asymmetric and often unstable decision criteria when choosing an appropriate response to a task at hand. There are, however, genuinely bias-free discrimination tasks, involving n-alternative, typically two-alternative, forced choice tasks (2afc tasks) (see also Phillips 2015b, for discussion). These tasks induce a symmetric criterion in subjects, thus guaranteeing that their performance is not skewed towards one of the options. Responses given in such forced choice tasks, provided the experiment is carefully set up to prevent changes in criteria due to different motivational structures, are therefore non-biased. Nonetheless, these responses are still *mediated* by decision processes of the subject: any response, according to SDT is the product of (perceptual) sensitivity and decision criterion. But being mediated does not mean being biased.

Thus, if there is bad news for subjective measures of consciousness in particular, it does not simply fall out of the assumption at the heart of SDT, namely that all subjective reports are the products of perceptual sensitivity and decision processes. This assumption does not yet make for a problem of response bias for subjective measures.

In the case of *measures of perceptual sensitivity*, recall, the concern raised by response bias is that one cannot simply read perceptual sensitivity off the average rate of subjects' accurate positive responses to the stimulus. The latter is not automatically an appropriate measurement of the former because it demonstrably includes the effect of subjects' decision criteria. To get an accurate estimate of perceptual sensitivity, this concern can be addressed by either using non-biased tasks (i.e. especially 2afc tasks), or, by taking biased responses and calculating perceptual sensitivity from the relationship between accurate positive responses and false positive responses, using a ROC curve analysis.

When it comes to *measures of consciousness*, if perceptual sensitivity is taken to be a measurement of consciousness (as, e.g. Eriksen proposes), then the concern about response bias is the same as

in the above case of measures of perceptual sensitivity, and so is the solution. But, if subjective reports (or, better, thresholds indicated by such reports) are taken to be the appropriate measurement of consciousness, then the problem that the presence of response bias poses is different. It is not that we have any direct evidence of response bias as we do in the perceptual sensitivity case. And, as I pointed out earlier, just because reports are mediated by decision criteria does not on its own provide a reason to think that they are biased. Rather, the evidence is indirect: it is simply implausible that consciousness fluctuates with response thresholds when the latter can be manipulated so easily and by factors that seem unlikely to be causally relevant to perceptual consciousness. So, it seems plausible that reports are contaminated by response bias. This is the problem of response bias for subjective measures of consciousness.

Is this problem intractable? We cannot use the same statistical approach to filter out bias that we use in the case of measuring perceptual sensitivity. The approach must be different: it must focus on making it less plausible that a given set of reports is contaminated by response bias, thereby reducing the implausibility that fluctuations in subjective threshold in a given case mark fluctuations in consciousness. To do so we need to know details about subjective measures: to even begin to distinguish which kinds of fluctuations might be due to consciousness itself and which kind might be due to response biases of various sorts, we need to understand the kinds of subjective access at work in subjective measures. Only once we have an understanding of the nature of subjective access, can we assess the prospects for eliminating or significantly reducing bias in the data provided by subjective reports.

As I argue in the remainder of the chapter, contemporary debate about subjective measures of consciousness does not provide such an understanding. The problem of response bias for subjective measures is therefore currently less tractable than it might be, since we are not in a position to adequately assess whether a given set of response-based data is likely infected by bias. I concentrate on three aspects of the debate in need of clarification. I first soften the ground in section 5, by showing that the distinction between subjective and objective measures is standardly not clearly drawn. In section 6, I discuss subjective measures directly, showing that the notion of subjective access at the heart of them is typically not worked out. I develop this point in section 7, by investigating the role (if any) of introspection in extant subjective measures.

5. The distinction between objective and subjective measures of consciousness

Although the distinction between objective and subjective measures of consciousness structures the overall area, it is in fact not easy to see what it comes to, based on how it is drawn by participants in the debate. The term ‘measures’ might refer to methods or to measurements. The latter, in turn, might also be understood in different ways: specific readings obtained on an occasion of measuring versus the general measurement unit. The objectivity or subjectivity of a given measure thus might depend on aspects of the method of measuring, or on aspects of the measurement. Extant characterisations of the distinction are typically brief and not explicit on this point, the presumption being that the basic idea of the distinction is generally understood. To pick a representative example, Seth et al. (2008) introduce the two types of measures as follows:

> “’Objective measures’ assume the ability to choose accurately under forced choice conditions as indicating a conscious mental state. .... ‘Subjective measures’ require subjects to report their

mental states."

On the objective side, characterizations often emphasise forced choice paradigms in classifying objective measures of consciousness. In the quote above, the focus seems to be on *measurements*, namely that objective measurements of consciousness are measurements of perceptual sensitivity, as defined by SDT. They are response-bias-free measurements of task performance in relation to stimulus intensity. In turn, for a *method* to count as objective, it requires a procedure to filter out any response bias skewing the data ('to eliminate bias' (Lau 2008)). Objective methods thus involve either the use of detection tasks that yield bias-free, accurate responses (as, e.g. 2afc tasks) to determine perceptual sensitivity, or use of more complex statistical methods, by submitting responses from biased perceptual detection tasks to a ROC curve analysis.

Another frequent suggestion, however, is that the objectivity of measures of consciousness consists mainly in the public observability of the data produced (Timmermans and Cleeremans 2016: 22). Again, the focus seems to be on measurements, namely that an objective measurement is publicly observable. An objective method is then a way of acquiring such publicly observable data. Public observability is typically associated with other general conditions, such as the need for replicability of results, systematic variation, etc. In that way, objective measures are often held to be scientifically valid measures more generally (see discussion in Overgaard 2016: 9).

Note that if the objectivity of a method consists in bias-free measuring, i.e. producing a measurement that is not influenced by any response bias, then it seems that in principle – should one be able to eliminate or filter out response-bias in this case – so-called subjective methods might count as objective, too. If the objectivity of a method consists in producing publicly observable data, i.e. measurements that are publicly available (replicable, systematically variable, etc.) then this does perhaps rule out subjective methods. On both points, much depends on what the latter exactly are, of course. We will turn to this question in a moment. However, at this point it is already clear that different formulations of the objectivity of measures (freedom from response bias, public observability) draw the line between subjective and objective measures differently.

Let us now turn to the other side of the distinction, subjective measures. The single most mentioned feature in characterisations of subjective measures is that they involve 'subjective reports'. A subjective *method* of consciousness gathers data about consciousness by getting subjects to report something relevant to consciousness. The *measurement* of a subjective measure is report. It is unclear, though, what makes for *subjective* reports. Is it that they (i) are acquired via a specific method, or (ii) have a specific type of content?

Sometimes subjective measures are characterised merely by appeal to the kind of report provided by subjects, i.e. in terms of subjective measurement:

> "The undeniable merit of subjective measures is their face validity: they are reports about the conscious experience itself, rather than reports about something else, that is the stimulus." (Lau 2008: 253-4)

This will not do, however, at least not without qualification. Distinguish between 'content-subjective-report' and 'content-objective report'. When asked to report on what they perceive, a subject might say:

1) ‘I see that there is a red triangle.’
2) ‘There is a red triangle.’

1) has a subjective content in the sense that it is about a mental feature, a seeing, i.e. a visual experience as of a red triangle. 2) has an objective content in the sense that it is about a worldly scene before one. Shea (2012: 310) notes, in ordinary talk 1) and 2) might be used interchangeably. Each might be held expressive of either, a claim about the world, or a claim about one’s experience of the world. He further points out, experimental psychologists typically take content-objective-reports to be evidence for relevant experiential states of the subject. Presumably, this is because they hold that such reports are expressive of the latter (see also Chirimuuta 2014). In addition, both kinds of report can be made verbally and non-verbally (e.g. button-pressing).

Lau’s characterisation, then, does not properly capture the distinction between types of measures. Another frequent characterisation implies that the subjectivity of subjective measures is a matter of method. The key feature here is that the experimenter instructs subjects in relevant ways to report their mental states.

> “In the case of a subjective measure, participants are asked specifically to respond according to their own internal state of awareness.” (Tunney and Shanks 2003: 1061)

Thus, the measurements - the reports - are subjective because of how they have been elicited by the experimenter from the subject. This still leaves room for different conceptions. The experimenter might ask subjects to:

i. report something about their experiences;
ii. report something about experience by accessing their experience in a specific (e.g., first-person) manner;
iii. report something, with the experimenter’s unstated expectation that in complying with this request subjects access their experiences in a specific manner).

On any of these three conceptions, there is room for taking *prima facie* objective-content-reports to be a species of subjective report. Having been gathered via a subjective method would mean that one can interpret such reports to be expressive of subjective-content-reports as necessary. But the three conceptions potentially offer significantly different answers to what makes reports subjective and hence, to what is characteristic of subjective measures. The first one (i.) does not provide much substance to the idea of a subjective measure and it is, at least without further details, consistent with the characterizations of objective measures discussed earlier. In (ii.) and (iii.), though, the idea is that by eliciting the report in the manner that they do, the experimenter gets the subject to access their experience in a specific way. This access is key to a substantial conception of subjective measures. Let us now turn to the main contemporary types of subjective measures and look at the notion of subjective access they employ.

6. Subjective measures

In the last twenty years, subjective measures of consciousness have seen a revival in the science

of consciousness (see, e.g. Jack and Roepstorff 2003, 2004). They rely on gathering data about the presence or absence of consciousness via subjective reports. The latter are typically made in relation to a primary task, involving objectively measurable performances (e.g. perceptual discrimination or identification, or, in the case of artificial grammar learning, judgements of grammaticality). Correlation between accuracy of primary task performance and various aspects of subjective reports form the basis for conclusions drawn about consciousness (Tunney and Shanks 2003). The main subjective measures of consciousness are Confidence Ratings (CR) and the Perceptual Awareness Scale (PAS), distinguished in terms of the type of subjective report they use.[5] Further distinctions among approaches can be made in terms of the kind of statistical analysis used to evaluate the data obtained via subjective reports. Here, I will focus on the first kind of difference (for discussion of different kinds of data analysis, see, e.g. Fleming and Lau 2014, Norman and Price 2016).

In recent years, proponents of different subjective measures have compared and contrasted the adequacy of these rival methods, but their studies have not yet produced consensus about which subjective measure produces the overall best results (see, e.g. Dienes and Seth 2010, Overgaard and Sandberg 2012, Sandberg et al. 2010, 2013, Wierzchón et al. 2012, Zehetleitner and Rausch 2013).

6.1 Confidence ratings

One experimental line of inquiry uses so-called 'confidence ratings' as an index for consciousness. Confidence ratings are subjects' reports about their own confidence concerning some state of affairs obtaining. In consciousness research, they concern subjects' perception of a given stimulus, or the accuracy of subjects' responses in a given primary task. The primary tasks - called 'type 1 tasks' - are perceptual discrimination or identification tasks.[6] These form the object of so-called 'type 2 tasks', i.e. assessment tasks where the subject has to judge and report on their confidence in how successful they were in a given type 1 task.

Researchers analyse correlations between performance in type 1 and type 2 tasks and draw conclusions about whether performance on type 1 tasks was influenced by consciously held information available in task 1. In particular, they tend to employ one of two criteria in their data

[5] There are other measures, such as Post-Decision Wagering (PDW) and No-Loss Gambling (NG) which are not easily classed as subjective or objective (e.g. Persaud et. al 2007). These are behavioural measures involving betting on something related to stimulus perception. I will concentrate on CR and PAS, since PDW and NG, on to the view that they are subjective measures, are confidence-related measures and similar to CR in most respects that are of interest here.

[6] Confidence ratings have more recently been used in research on artificial grammar learning (AGL). The stated aim is to probe the influence of conscious and unconscious knowledge in implicit learning - 'the phenomenology of the application of knowledge resulting from implicit learning' (Dienes et al. 2010: 685). It is not straightforward how to think of the notion of phenomenal consciousness at issue in this research paradigm, in relation to the one associated with sensory perception. I am concentrating on perceptual consciousness here.

analysis: the non-zero correlation criterion and the guessing criterion (Dienes et al. 1995). The former looks at the relationship between confidence ratings and accurate (type 1) performance: if there is no significant positive relationship between confidence rating and performance, the performance is held to be influenced by unconscious perception. The latter looks at the relationship between low-confidence ratings, where subjects claim to be merely guessing, and accurate (type 1) performance: if accurate type 1 performance is above-chance in the case of low confidence ratings, the performance is held to be influenced by unconscious perception.

Confidence ratings have long been held to be good indicators of conscious episodes. According to Snodgrass et al. (2009: 563), 'confidence ratings are, plausibly, about as basic an indication of subjective experience as there is'. Traditionally, confidence ratings are used as indicators of perceptual experience – i.e. of phenomenal consciousness involved in (mainly visual) perception. In their classic paper, Peirce and Jastrow (1885), assume that there is a connection between subjects reporting confidence in their own perceptual discriminations and subjects being consciously aware of the stimulus, supported by their observation that overall high confidence tracks accuracy (for relevant discussion, see Fleming and Lau 2014). The seminal work by Merikle and Cheesman mentioned above also uses confidence ratings as their main gauge of the subjective threshold of perceptual awareness.

So, what are confidence ratings? Despite the widespread conviction about their good evidential status with respect to experience, it is unclear exactly what proponents of CR have in mind by confidence ratings and how they understand the basic relation between such ratings and any target experience. Here is a representative quote introducing confidence ratings:

> "A *confidence rating* is a self-report rating of one's confidence in a judgment or decision, usually given retrospectively after the judgement has been made. It involves assessing the validity of an assertion or a prediction …. Confidence judgements are metacognitive, in that they involve 'cognition about one's own cognition' (Metcalfe 2000). They can be seen as belonging to the subcategory of *metacognitive experiences* which reflect 'what the person is aware of and what she or he feels when coming across a task and processing the information related to it' (Efklides 2008, p.279)." (Norman and Price 2016: 159)

Note the following basic features of confidence ratings. Firstly, confidence reports are (verbal or non-verbal) reports that express a subject's confidence judgement. When I say 'I am highly confident that I saw a right-oriented grid', I am reporting my judgement that I am highly confident that I saw a right-oriented grid. The term 'confidence rating' is best reserved for the confidence report, though in the literature it is often used ambiguously for both report and judgement.[7]

Secondly, the confidence judgements are judgements about the subject's own level of confidence concerning a certain state of affairs. In CR, these states of affairs are either the subject's perception

[7] The distinction between confidence ratings (reports) and confidence judgements expressed by the reports does not play a significant role in the issues I am raising in this chapter. But the basic distinction between the cognitive upshot of a given kind of subjective access and the report expressing this upshot does matter to various questions about subjective measures not discussed here, e.g., concerns about the suitability of numerical or other kinds of scales to articulate subjective judgements, or about flawed articulation more generally.

of a stimulus, or the accuracy of their performance in a perception-based task. Confidence judgements are *not* the subject's states of confidence concerning these states of affairs themselves. They are the subject's take on their own level of confidence, they represent what the subject thinks their own level of confidence in this case is. Judging that one is highly confident in having made the right choice is not the same as being highly confident that one has made the right choice. Compare: judging that one is in pain, even sincerely, is not the same as being in pain. Just like in the case of pain, it may be hard to imagine that confidence judgements could be mistaken and, if so, such mistakes might be only possible in borderline cases. But none of that can be taken for granted without being spelled out further.

Thirdly, confidence judgements are not directly about the presence or absence of conscious perception. Indeed, according some proponents, this is one of the best features of CR: confidence ratings are unlike traditional introspective methods in that they do not aim to report experience directly, thereby avoiding the problems that are inherent in such introspective directness (e.g. Tunney and Shanks 2003: 1061). Instead, proponents of CR use terms like 'reflects' or 'indicates' to talk about the relation between confidence ratings (or judgements) and the target phenomenal experience involved in a given primary task. The idea is to 'ask participants to report their phenomenal states by means of confidence ratings' (Tunney and Shanks 2003: 1061), where this is not an instruction to subjects but rather a description of what confidence ratings can be held to be indirectly reporting.

There is a surprising lack of detail on how confidence judgements are revealing of any perceptual consciousness involved in type 1 performances. What is this reflection relation between confidence judgements and relevant phenomenal experiences, such that the former can serve as good evidence for the presence or absence of the latter? One option would be that reflection is a two-step process. The confidence judgement is about a state of confidence and formed on the basis of that state of confidence. In turn, the latter itself is formed in a way that is responsive in part to the presence or absence of relevant phenomenal experience. Thus, the relation that underwrites reflection has two nodes, confidence judgement to state of confidence, state of confidence to relevant phenomenal experience. Another option would be that reflection is a one-step process, in virtue of a relation between confidence judgement and phenomenal experience. The subject's take on their own states of confidence is formed in part directly on the basis of the presence or absence of relevant phenomenal experience. The second option seems more in line with how proponents of CR talk about confidence ratings, but this may be due to the fact that they often conflate confidence judgements with states of confidence.[8]

An added difficulty in reconstructing the overall picture is the frequent reference to feelings of confidence, or feelings of knowing. Confidence ratings are said to either reveal, or be based, on such feelings (e.g. Norman and Price 2016: 165). There does not seem to be any developed story about how these experiences interact with the target experiential episodes involved in type 1 performances of perceptual tasks. *Prima facie*, feelings of confidence or knowledge are phenomenal experiences in their own right and it is not obvious how they relate to perceptual

[8] If taken at face value, Tunney and Shanks (2003: 1063) seem to hold that the reflection relation is a two- step affair, the middle step involving a belief that one is correct in one's discrimination. 'When participants are aware of the knowledge used to make a discrimination, they presumably believe themselves to be correct and should respond with high confidence'.

experiences. They also seem distinct from states of confidence, at least according to standard conceptions of those as belief-like propositional attitudes that come in degrees, or beliefs about degreed propositional contents. It is hard to fit feelings of confidence, or knowledge into the narrative above in a way that facilitates the evidential relation confidence judgements supposedly bear to the absence or presence of target phenomenal experiences. Comments such as 'the subjective feeling of knowing that one has perceived a stimulus often accompanies our visual experience and should be considered an important aspect of visual consciousness' (Samaha et al. 2016) are not all that helpful in fleshing out the picture.

The heart of the CR programme seems to rest on a brute intuition about an epistemic connection between confidence ratings and conscious character of target perceptions: when detection of a stimulus involves conscious perception of the stimulus, high confidence judgements concerning detection are made in response to these conscious episodes.

> "These criteria of awareness assume that if participants are aware of the knowledge they use to classify items, they should be more confident in correct than in incorrect decisions. It follows that the information used to make confidence ratings should be the same as that used to classify items." (Tunney 2005: 368)

It is assumed that where confidence judgements track accuracy in type 1 tasks, a subject's take on her own confidence concerning accurate perceptual detection or discrimination of a stimulus flows from the subject's conscious perception of that stimulus, such that we can take the former as an indication of the latter. Not much more is said about this crucial basing relation providing the nexus between confidence judgements and conscious perception.

However, without further articulation of the connection between confidence judgements and target conscious episodes which underlies the use of confidence ratings as measurements for the presence or absence of consciousness, it is hard to see how to make progress with the problem of response bias for subjective measures of consciousness. Recall, response bias is a 'measure of the participant's tilt' towards one of the response options. One of the key insights of SDT is that decisions to respond are made in light of goals (e.g. to maximize correct positive responses, or to avoid punishment for incorrect positive responses, etc.) and that these can have a significant part in shaping the responses. In the case of subjective measures, the approach I suggested we need to take, is to identify and understand relevant kinds of 'tilt' subjects might be susceptible to when providing subjective reports. To do so – especially with an eye to perhaps substantially reducing, compensating, or filtering out response bias – we would need to have a clearer picture of what subjects are engaged in, what they are doing, when they subjectively access the relevant data.

Relatedly, it is difficult to see what exactly the subjectivity of CR (as a subjective measure) consists in, since we do not have a filled-out picture of the method of acquisition of the measurement data, i.e. of the subjective access used by subjects in providing the ratings. Confidence ratings are subjective reports at least in sense i.) above, in that they are about the subject's mental states – though these are not the target conscious episodes. Confidence ratings are also subjective in sense iii.) above, in that there is the experimenters' background expectation that in producing them subjects access their experiences in a specific manner. But what we need and do not have, is an account of the nature of the subjective access at work in generating the data about conscious experience.

6.2 Awareness ratings

Awareness ratings are the main source of data about conscious perception in a subjective measure called 'perceptual awareness scale' (PAS) (Ramsøy and Overgaard 2004). PAS shares with CR the basic thought that conclusions about consciousness can be drawn based on a correlation between subjective ratings (in this case awareness ratings) and performance in a type 1 task (see Sandberg and Overgaard 2016: 190-192 for a brief overview of statistical analyses of PAS data). Thus, experiments using PAS standardly involve a type 1 perceptual detection task, and a type 2 rating task.

Awareness ratings are reports that are meant to be directly about experience: they result from subjects being asked to 'report on their experiences directly and allowing them to do this on a scale created either by themselves or other participants presented with similar stimuli' (Sandberg et al. 2013: 806). In contrast to CR, the type 2 tasks involved in PAS are not about the performances in type 1 tasks, but instead are directed at any perceptual experiences involved in type 1 tasks.[9]

Again, let us distinguish between report and the judgement expressed by a report, reserving the term 'awareness rating' to refer to awareness reports. Awareness judgements are about the conscious character of the target perceptual episode. The stated aim by proponents of PAS, is for the content of awareness ratings to overlap with the phenomenal content of the perceptual episodes involved in type 1 tasks. [10] PAS seeks to provide an experimental setting 'where reports stand in a "1:1 relationship" with the other relevant inner states' (Sandberg and Overgaard 2016: 182). This means not only that for any awareness judgement, there is a corresponding awareness episode (or lack thereof, in case the awareness judgement is a negative one). The ambition is also that there is a certain descriptive accuracy, so that the content of the awareness judgement (and thus the awareness rating expressing it) adequately captures the phenomenal character of the corresponding episode in relevant respects.

In particular, awareness ratings aim at the key feature of clarity of a given perceptual experience. According to them, conscious perceptual content comes in 'different degrees of clarity' and they base their claim both, on empirical work demonstrating preserved residual awareness in blindsight patients (Zeki and ffytche 1998), as well as on simple contemplation of ordinary cases of perceptual experiences. The feature of clarity thus is meant to capture the level or strength of a subject's perceptual awareness of a given stimulus. Low clarity represents partial, weak, or degreed awareness of a stimulus.[11] Ramsøy and Overgaard take the graded nature of experience to

[9] PAS tends to be used in investigations of visual and auditory perception, but not in investigations of conscious grammar knowledge.

[10] I use 'overlap' to remain neutral on the extent to which phenomenal character can be captured in judgements and, furthermore, be articulated in reports. Thus, ideally, awareness judgements have the same content as the target perceptual episode. Less ideally, some details are lost or added in the process of judging and articulating.

[11] I am emphasising this because from the point of view of certain debates in philosophy of perception, clarity might be thought to refer to the content of phenomenal character in a different way. An experience might present objects or properties clearly or not very clearly while the overall conscious episode is fully conscious and not faint. For instance, the phenomenal character of a perceptual episode as of a cluttered beach might have a precise content in that one

motivate a non-dichotomous, multiple step scale of awareness ratings in experiments designed to test for subliminal perception.

One of the key ideas underpinning PAS is that the multiple-step response scale is developed in co-operation with participants in the experiments, specifically to ensure the right number of steps and an adequate description of the clarity of awareness captured by a given step. Application of PAS therefore begins with certain amount of training administered to participating subjects, followed by an active development phase. Ramsøy and Overgaard (2004)'s instructions for constructing the PAS steps, for instance, ask subjects to report the clarity of their experience by using a scale that is bounded on each side by 'no experience at all' and 'a clear image', but which can otherwise have as many steps as subjects find useful. In addition, every experiment begins with explaining and extensively discussing the meaning of the different scale steps with participants, occasionally even interrupting trials to check with subjects why they used a given scale step in their response. If the experiments and general setup are similar enough, a PAS developed in one investigation may be used in another, or it may be used as a starting point to be tweaked in further experiments.

In part, this aspect of PAS is designed to alleviate concerns about the scientific validity of awareness ratings, particularly concerns about response bias and about the lack of inter-measurement checks (Sandberg and Overgaard 2016). Individual differences and disagreements due to individual bias are claimed to be minimised because experimenters and participants jointly develop the scale, thereby establishing common ground. Furthermore, developing the response scale in this way is meant to facilitate overlap between awareness rating and the phenomenal content of the target perceptual episode – getting as close as possible to a 1:1 relationship between them.

In contrast to CR, PAS also seems to provide a more detailed conception of how awareness judgements epistemically relate to the target conscious perceptions: awareness judgements are directly about perceptual experience and subjects arrive at the former by introspecting the latter. This seems to yield a more precise answer to what the subjectivity of the PAS consists in. Awareness ratings are subjective reports in the sense of ii) above because they express judgements that are about experience arrived at via introspective access to experience. But as I argue in the next section, this impression is misleading. While proponents of PAS use the term 'introspection' in picking out the type of subjective access at work in their method, they offer little elucidation of what introspection is. Again, without further details about the nature of the subjective access employed in generating the report data – in this case introspective access – we cannot get a precise understanding of the kinds of bias this access is susceptible to. This means that we cannot make progress on the problem of response bias for subjective measures of consciousness. It is, for instance, entirely unclear how the consensus building aspect of PAS is meant to interact with the individual's subjective access exploited in the experiments.

---

is visually aware of objects on the beach in sharp detail, or it might have a more fuzzy content in that one cannot easily differentiate different object because things look blurry. But either of those visual experiences could still be clear in Ramsøy and Overgaard's: they could be fully consciously present. It is just that what is presented might be blurred or not defined.

7. Subjective Access and Introspection

Proponents of CR and PAS presuppose considerable familiarity and competence with the kind of first-person access required by their methods. Yet they do not put forward a substantial view of the nature of such access. This combination shows in the sparse instructions given to subjects in experiments, as well as in the lack of elucidation given of access-denoting terms in their discussions within the research community.

On the instruction side, proponents of subjective measures expect experimental subjects to be mostly competent in complying with instructions to select and use the required kind of subjective access.

> "(P)articipants usually do what they are asked, so when one asks them about experiences, one should expect that they report their experiences, and when one asks them about their confidence, one should expect that they report their confidence (and not their experience)." (Sandberg and Overgaard 2016: 187)

But asking them to so typically does not involve a lot of detailed direction or guidelines. In standard CR experiments, subjects are simply asked to assess how confident they are in their own type 1 task responses, and they are told to use various kinds of response scales to express their confidence (see, e.g. Norman and Price 2016, Samaha et al. 2016, Tunney 2005, Zehetleitner and Rausch 2013). The practice clearly presupposes that different participants will be making their assessments in the same manner and that the cognitive process involved in each case draws on the presence of the target conscious perception.

In PAS experiments, one might expect that the kind of training involved in preparing subjects for, and guiding them through, developing an awareness scale involves specific explanations and instructions of what to do in order to introspect their experience. However, at least as far as typical published descriptions of PAS procedure are concerned, instructions of this kind, i.e. about what to do to introspect, are scant. The procedure takes for granted that subjects know what the basic kind of subjective access is that they need to employ when they are 'asked to report (or guess) all three stimulus properties (shape, color, and position) and then report the clarity of each property using a scale that they developed themselves' (Sandberg and Overgaard 2016: 182). This is not to say that there are no instructions offered – quite the contrary, PAS involves an extended preparatory phase consisting of discussion with participants about the kind of verdicts they might render when they introspect experience (Sandberg and Overgaard 2016: 188). But these instructions and the attending training effort are about getting subjects to understand, and help configure, the steps of the scale used in introspection, and are focused on the content of the different awareness ratings they are making. They do not concern the kind of subjective access itself that is supplying these verdicts.

Yet, invariably, proponents of subjective methods insist that there are different forms of subjective access at work in their respective methods. For instance, the PAS was initially prompted by the thought that, although subjective measures seem needed for an investigation of conscious perception, confidence ratings cannot automatically be taken to measure conscious awareness because they are not sourced in a sufficiently direct way (Ramsøy and Overgaard 2004). There is broad agreement that PAS involves introspection, whereas CR does not, or at least not in the same,

straightforward manner that PAS does, with typically suggestive but not very detailed remarks about the difference (e.g. Overgaard and Sandberg 2012: 1290).

What makes it difficult to understand this difference is that practitioners neither offer a worked-out view about what introspection is, nor do they use consistent terminology (compare e.g. Overgaard and Sandberg 2012, Wierzchón et al.: 2014, and Fleming and Lau 2014). One might be forgiven for suspecting that there is no substantial notion of introspection – or subjective access – in circulation in this area.

Such a suspicion is particularly apt when it comes to CR. There is no consensus in the literature on subjective measures whether confidence judgements are introspective in some significant sense. Some hold confidence judgements (and hence ratings) to be introspective, while others explicitly distinguish them from introspective judgements. Among the former, there is a further difference in how much emphasis they place on such a classification. Many seem to be using 'introspective' interchangeably with 'metacognitive' or even 'subjective', and are best understood as not putting forward a positive view about any special subcategory of metacognition or introspection (e.g., Frith and Lau 2006, Lau 2008). Others, however, seem more committed:

> "The central premise here is that confidence ratings reflect subjective experience (or phenomenal consciousness; Block 1995), because they are by definition a variety of introspective reports about our mental processes." (Norman and Price 2016: 160)

On such a view, the crucial relation which supports the evidential link between confidence ratings and conscious perception is introspective. Next to no detail is offered, though, on what is it about the subject's own assessment of the accuracy of their own type 1 task performance that makes it count as introspective - or, for that matter, even as first-personal. Given that subjects could make such accuracy assessments in all sorts of ways, e.g. by induction from similar past performance, this is not an idle point.

Proponents of PAS, as noted in the previous section, are clear that the awareness ratings at issue in PAS express awareness judgements that are gotten by introspection.

> "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)." (Overgaard and Sandberg 2012: 1291-2)

This is meant to ensure that awareness ratings are directly about experience. Overall, PAS aims to provide an introspective, scaled measure of conscious content that is more directly about experience and more sensitive to its graded nature than CR. However, here too, the notion of introspection itself is not elucidated in serious detail.

Introspection is sometimes held to be *by definition* a process which involves a judgement (or similar) that is about experience - introspection is said to bear 'a necessary link to conscious experience' (Overgaard and Sandberg 2012: 1288). It thus is the target of introspection - experience - that differentiates introspective from non-introspective types of metacognition, in the first instance. Standard expressions of this view fluctuate between making a claim concerning the proper inputs of introspection (that it must be an experience), and making a claim concerning the

proper outputs of introspection (that the content of introspective judgements or episodes must be about an experience). As an example of the latter, Overgaard and Sandberg (2012) insist that 'it should be clear that introspective report by definition is only about what is experienced and not about cognitive processes'. Although the two claims are independent, there seems to be an expectation that they go together. The thought is that introspection takes experiences as input and has as outputs states or episodes the content of which is about experience. When things work well, the content of the output is expected to match the content of the experience in some relevant sense.

Sometimes this characterisation of introspection is further filled out by an indication about the specific mental process involved in introspecting. For instance, Ramsøy and Overgaard (2004) gloss introspection as 'a state in which one directs attention towards one's own experiences'. Here it seems introspection is held to be a metacognitive process that essentially involves attention or an observation-like relation to one's own conscious states and episodes (see also Overgaard and Sandberg 2012: 1288). But this initial gloss on introspection as attention to experience is meant to be intuitive and pre-theoretic, not a substantial proposal. It is deliberately permissive with respect to different accounts of the nature of the introspective attitude and its relationship to the conscious states and episodes that form the attentional target of introspection. For instance, the gloss is taken to be neutral on the question of on-line, concurrent processing versus retrospective accounts of introspection. Nor is it meant to determine whether introspectively attending to experience causally affects the latter in some way.

Proponents of CR and PAS sometimes speculate about the underlying cognitive or neural architecture of subjective access. Proposals tend to be based on empirical evidence about dissociations between awareness ratings and confidence ratings, revealed by different sensitivities of the respective awareness and confidence scales in experiments investigating visual perception. A variety of studies have offered evidence for superior sensitivity of PAS, CR, and CR-related measures, respectively (see references at the beginning of section 5). As a consequence, there are several empirically motivated proposals for a neurally-based or computational difference between introspective (awareness) measures and confidence measures. These proposals agree that there is such a difference, but they disagree about what the empirical difference consists in, about the details of the neural or computational mechanisms manifesting it, and about the precise empirical evidence for it.

Overgaard and Sandberg (2012), for instance, suggest on the basis of differential sensitivity of PAS and CR, that one might think of introspective judgements as involving a simpler cognitive process than that driving confidence judgements. The latter draw on a more complicated process that includes some introspective processing, but also further processes that yield the kind of metacognitive insight into how accurate one's own discrimination or identification processes are. Zehetleitner and Rausch (2013) argue that the systematic differences in sensitivity between awareness and confidence scales are best explained by taking the judgements involved to be sensitive to different neural events. In turn, they reason that the respective ratings are measures of different cognitive processes, i.e. that they track different functions in the cognitive economy. They suggest that awareness ratings track 'the strength or the quality of the internal signals that form part of the sensory data', while confidence ratings track 'those internal signals that are involved in

the decision to make a response' (1423).[12]

Such views about the underlying neural or cognitive nature of introspective processes are, although motivated by empirical evidence, highly speculative. Their proponents are usually happy to admit as much (Overgaard and Mogensen 2017, Overgaard and Sandberg 2012, Zehetleitner and Rausch 2013 all include serious hedges, for instance). The speculations do seem worthwhile, I suspect, because they appear to forge a path towards obtaining empirically grounded accounts of introspection and metacognition, along the same lines as vision science has done for what we ordinarily might refer to as 'sight' - namely, of visual perception as a form of perception more generally. This is an enticing prospect. However, for all its promise, such a prospect still faces enormous hurdles.

The main hurdle is what I call 'the problem of the starting point'. In short, it is usually not clear what the initial psychological phenomenon is in the first place, with which, or about which we are theorising. There is no basic, core common-sense take on introspection, or on kinds of first-person access, that constrains the beginning of our enquiries in the same way that common-sense understanding of sight, audition, etc. does. We ordinarily use the term 'introspection' in all sorts of ways and every day talk and practice contains a wide array of ideas of first-person-sourced awareness. The notion of introspection is much more like the common-sense notion of thought, i.e. vague, amorphous, and fluid, ranging over active, deliberative inference, quasi-perceptual attention, retrospective episodic memory, passive self-awareness, etc. It does not support the idea that there is a theoretically significant psychological kind or group of mental capacities, or that available empirical evidence gathered in terms of it supports theorizing about neural and cognitive architecture.

As I have shown above, the crucial notions of introspection and subjective access at work in the key assumption characterizing the psychological literature on subjective measures, namely that we have certain first-person ways to access aspects of our conscious lives, are never fully articulated but instead merely gestured towards, relying on a common-sense understanding of them. Insofar as the notion of introspection, for instance, is given any elucidation at all, it tends to be dealt with in one or two generic sentences. Fashioning an introspective (or subjective) method out of that gesture means that a fair amount of inchoate common-sense intuition about our first-person access is placed at the heart of it.

Proponents of CR and PAS agree that all subjective methods involve metacognition of some kind, but this does not offer much of a common starting point either.[13] Metacognition is a more general category than introspection, covering any mental state or episode about another mental state, or episode, or process. It is a higher-order state or episode in the sense that it has a first-order, or

[12] Along a different dimension of the nature of introspection, Overgaard et al. (2006) take themselves to provide empirical evidence against the view that introspection is always retrospection.

[13] Proponents of PAS sometimes distinguish introspection from metacognition. Since they are also sometimes happy to think of introspection as a subclass of metacognition, I am going to attribute the latter view to them and adjust where necessary, chalking discrepancies up to terminological infelicities on their part.

lower-order mental state, episode or process as its object.[14] The class of metacognitions is thus large and motley: 'metacognition', as Fleming et al. (2012) put it, is 'an umbrella term'. All sorts of cognitions count as metacognitive cognitions, ranging from person-level judgements that one has a headache, to low-level states of early visual processing to the effect that some visual representation of an edge as oriented to the right is likely to be correct. There is no restriction, that is, on either the metacognitive output or input on that score (i.e. whether they are attributable to an individual or a subsystem), other than that they must be cognitive properties of a given psychological system. Moreover, the broad conception of metacognition does not place constraints on the metacognitive process, i.e. on how the output is produced. Thus, early visual automatic processing, conscious deliberation, attention, etc. are all permissible metacognitive processes. There is in principle no restriction to acquiring a metacognitive judgement, say, via testimony by others, or by looking in a mirror and inferring from the expression on one's face.[15]

8. Conclusion

The discussion in this chapter shows that current-day subjective measures of consciousness do not involve worked-out conceptions of introspection, or of other relevant types of subjective access. We have also seen fairly careless uses of the term 'introspection', even by those who hold that introspective ratings are importantly different from other subjective ratings, where it sometimes covers a specific metacognitive capacity only, and sometimes it is a catch-all term for any capacity underlying a variety of subjective measures. In light of this, we cannot even begin to address the problem of response bias for contemporary subjective measures of consciousness. We do not sufficiently understand the kind of subjective access that is employed in these measures. There is then no hope of trying to figure out whether response bias is inextricably and debilitatingly inherent in the subjective method itself, or whether it can be minimized or filtered out by ingenious experiment or data analysis.

In the introduction, I noted Irvine (2012: 629)'s observation that critiques of subjective measures

[14] This is perhaps the most permissive characterisation of metacognition, i.e. it is the broadest understanding of metacognition as 'a cognition about cognition' (Fleming et al. 2012: 1280) Insofar as one holds that cognition is a representational phenomenon, metacognition is a system's representation of its own representations. It is not thereby guaranteed that the aspect of cognition represented by the metacognitive output is the causally efficacious input, nor that the aspect of cognition represented concerns the representational content of the represented cognitive state. For a conception of metacognition as meta-representation, see (Shea 2014: 315-6)). For other, narrower conceptions, see, e.g. Fleming (2017).

[15] Those working on metacognition are, it seems, mainly interested in processes that are in some sense more strictly part of our natural psychological system, and so testimonial relations and other complex inferential methods would not be considered fundamentally metacognitive. But metacognition is normally introduced in terms of a functional characterization, i.e. as serving to control and monitor behaviour, and this does not place much of a constraint on metacognitive process, inputs or outputs. Moreover, the relevant behaviour can be held to include the activity of cognitive processes, or it can be restricted to agent-level behaviour. Depending on the range of behaviours at issue, metacognition may be more or less wide a category. See (Fleming et al. 2012) for interesting discussion.

of consciousness are echoing well-known objections to early introspectionist psychology at the turn of the last century, thus indicating a lack of progress in this area of current-day science of consciousness. Indeed, when it comes to the problem of response bias, I think there is not merely stagnation, there is a setback. Early experimentalist psychologists, such as Wilhelm Wundt and Georg Elias Müller, had developed views about the role and nature of subjective access in their introspective methods, quite different from the caricature views attributed to so-called 'introspectionist psychologists' (Spener 2018). The debate about the problem of response bias for subjective measures of consciousness would be much enhanced by revisiting these older views because the latter provide at least a blueprint for how to address this problem by being clear about what subjective access consists in.

References

Block, N. (2008), 'Consciousness and cognitive access', in *Proceedings of the Aristotelian Society* 108: 289–317.

Chalmers, D. (2004), 'How can we construct a science of consciousness?', in Michael Gazzaniga, editor, *The Cognitive Neurosciences III*. MIT Press.

Cheesman, J. and Merikle, P.M. (1984), 'Priming with and without awareness' In *Perception and Psychophysics* 36: 387–395.

_____. (1986), 'Distinguishing conscious from unconscious processes', in *Canadian Journal of Psychology* 40.

Chirimuuta, M. (2014), 'Psychophysical methods and the evasion of introspection', in *Philosophy of Science* 81(5): 914–926.

Dienes, Z. and Seth, A. (2010), 'Measuring any conscious content versus measuring the relevant conscious content: Comment on Sandberg at al', in *Consciousness and Cognition*, 19(4): 1079–1080.

Dienes, Z. and Altmann, G.T.M. and Kwan, L. and Goode, A. (1995), 'Unconscious knowledge of artificial grammars is applied strategically'. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25: 1322–1338.

Dienes, Z. and Scott, R. B. and Seth, A. (2010), 'Subjective measures of implicit knowledge that go beyond confidence: Reply to Overgaard et al.', in *Consciousness and Cognition*, 19: 685-686.

Eriksen, C. W. (1956), 'An experimental analysis of subception', in *American Journal of Psychology*, 69: 625–634.

_____. (1960), 'Discrimination and learning without awareness: a methodological survey and evaluation', in *Psychological Review*, 67: 279–300.

Fleming, S. M. and Lau, H. (2014), 'How to measure metacognition', in *Frontiers in Human Neuroscience*, 8, 2014.

Fleming, S. M. and Dolan, R. J. and Frith, C. D. (2012), 'Metacognition: computation, biology and function', in *Philosophical Transactions of the Royal Society of Biological Sciences*, 367: 1280–1286.

Frith, C. D. and Lau, H. (2006), 'The problem of introspection', in *Consciousness and Cognition*, 15: 761–764.

Green, D. M. and Swets, J. A. (1966), *Signal Detection Theory and Psychophysics* (Wiley).

Holender, D. (1986), 'Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal', in *Behavioral and Brain Sciences*, 9: 1–66.

Irvine, E. (2012), 'Old problems with new measures in the science of consciousness', in *British Journal for the Philosophy of Science*, 63: 627–648.

_____. (2013), 'Measures of consciousness' in *Philosophy Compass*, 8(3): 285–297.

Jack, A. and Roepstorff, A. (2003), *Trusting the Subject? The Use of Introspective Evidence in Cognitive Science, Volume 1* (Imprint Academic).

Jack, A. and Roepstorff, A. (2004), *Trusting the Subject? Volume 2* (Imprint Academic).

Lau, H. (2008), 'Are we studying consciousness yet?' in L. Weiskrantz and M. Davies (eds.), *Frontiers of Consciousness* (Oxford University Press).

MacMillan, N. A. and Creelman, C. D. (1991), *Detection Theory: A User's Guide* (Cambridge University Press).

Marcel, A. (1980), 'Conscious and preconscious recognition of polysemous words: Locating the selective effect of prior verbal context', in R. S. Nickerson (ed.), *Attention and Performance VIII*, (Hillsdale, N.J.: Erlbaum), 435–457.

_____. (1983), 'Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes', in *Cognitive Psychology*, 15: 238–300.

Merikle, P. M. and Cheesman, J. (1986), 'Consciousness is a 'subjective' state', in *Behavioral and Brain Sciences*, 9: 42–13.

Norman, E. and Price, M. C. (2016), 'Measuring consciousness with confidence ratings', in M. Overgaard (ed.), *Behavioral Methods on Consciousness Research* (Oxford University Press), 159–180.

Overgaard, M. (2006), 'Introspection in science', in *Consciousness and Cognition*, 15: 629– 633.

_____. (2016), 'The challenge of measuring consciousness' in M. Overgaard (ed.), *Behavioural*

Methods in Consciousness Research (Oxford University Press).

Overgaard, M. and Mogensen, J. (2017), 'An integrative view on consciousness and introspection', in *Review of Philosophical Psychology*, 8: 129–141.

Overgaard, M. and Sandberg, K. (2012), 'Kinds of access: different methods for report reveal different kinds of metacognitive access', in *Philosophical Transactions of the Royal Society of Biological Sciences*, 367: 1287–1296.

Overgaard, M. and Koivisto, M and Sørensen, T. A. and Vangkilde, S. and Revonsuo, A. (2006), 'The electrophysiology of introspection', in *Consciousness and Cognition*, 15: 662-672.

Peirce, C. S. and Jastrow, J. (1885), 'On small differences in sensation', in *Memoirs of the National Academy of Sciences*, 3(73-83).

Persaud, N., McLeod, P., Cowey, A. (2007), 'Post-decision wagering objectively measures awareness', in *Nature Neuroscience*, 10, 257-261.

Phillips, I. (2015a), 'No watershed for overflow: Recent work on the richness of consciousness', in *Philosophical Psychology*:1–14.

_____. (2015b), 'Consciousness and criterion', in *Philosophy and Phenomenological Research*:1–33.

Ramsøy, T. Z. and Overgaard, M. (2004), 'Introspection and subliminal perception', in *Phenomenology and the Cognitive Sciences*, 3(1): 1–23.

Reber, A. S. (1967), 'Implicit learning of artificial grammars', in *Journal of Verbal Learning & Verbal Behavior*, 6: 855–863.

Samaha, J. and Barrett, J. J. and Sheldon, A. D. and LaRocque, J. J. and Postle, B. R. (2016), 'Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory', in *Frontiers in Psychology*, 7.

Sandberg, K. and Overgaard, M. (2016), 'Using the perceptual awareness scale (PAS)', in M. Overgaard (ed.), *Behavioral Methods on Consciousness Research* (Oxford University Press), 181–195.

Sandberg, K. and Timmermans, B. and Overgaard, M. and Cleeremans, A. (2010), 'Measuring consciousness: Is one measure better than the other?' *Consciousness and Cognition*, 19(4):1069–1078.

Sandberg, K. and Bibby, B. M. and Overgaard, M. (2013), 'Measuring and testing awareness of emotional face expressions', in *Consciousness and Cognition*, 22(3): 806– 809.

Schwitzgebel, E. (2011), *Perplexities of Consciousness* (MIT Press).

_____. (2013), 'Reply to Kriegel, Smithies and Spener', and in *Philosophical Studies*.

Anil Seth, A. and Dienes, Z. and Cleeremans, A. and Overgaard, M. and Pessoa, L. (2008), 'Measuring consciousness: relating behavioural and neurophysiological approaches', in *Trends in Cognitive Sciences*, 12: 314–321.

Shea, N. (2012), 'Methodological encounter with a phenomenal kind', in *Philosophy and Phenomenological Research*.

_____. (2014), 'Reward prediction error signals are meta-representational', in *Nous*, 48 (2): 314–341.

Snodgrass, M. and Lepisto, S. A. (2007), 'Access for what? reflective consciousness', in *Behavioral and Brain Sciences*, 30(5): 525–526.

Snodgrass, M. and Kalaida, N. and Winder, S. E. (2009), 'Access is mainly a second-order process: SDT models whether phenomenally (first-order) conscious states are accessed by reflectively (second-order) conscious processes', in *Consciousness and Cognition*, 18(2): 561–564.

Spener, M. (2013), 'Moderate scepticism about introspection', in *Philosophical Studies*.

_____. (2015), 'Calibrating introspection', in *Philosophical Perspectives*.

_____. (2018), 'Introspecting in the Twentieth Century', in A. Kind (ed.) *Philosophy of Mind in the 20th Century* (Routledge).

Timmermans, B. and Cleeremans, A. (2006), 'How can we measure awareness? an overview of current methods', in M. Overgaard (ed.), *Behavioral Methods on Consciousness Research* (Oxford University Press).

Tunney, R. J. and Shanks, D. R. (2003), 'Subjective measures of awareness in implicit cognition', in *Memory and Cognition*, 31: 1060–1071.

Tunney, R. J. (2005), 'Sources of confidence judgements in implicit cognition', in *Psychonomic Bulletin & Review*, 12(2): 367–373.

Wierzchón, M. and Asanowicz, D. and Paulewicz, B. and Cleeremans, A. (2012), 'Subjective measures of consciousness in artificial grammar learning task', in *Consciousness and Cognition*, 21(3): 1141–1153.

Wierzchón, M. and Szczepanowski, R. and Anzulewicz, A. and Cleeremans, A. (2014), 'When a (precise) awareness measure became a (sketchy) introspective report', in *Consciousness and Cognition*, 26: 1–2.

Zehetleitner, M. and Rausch, M. (2013), 'Being confident without seeing: What subjective measures of visual consciousness are about', in *Attention Perception and Psychophysics*, 75 (7): 1406–1426.

Zeki, S. and ffytche, D. H. (1998), 'The Riddoch syndrome: Insights into the neurobiology of conscious vision', in *Brain*, 121: 25–45.