Who's afraid of response bias?

Megan A. K. Peters[1,*], Tony Ro[2] and Hakwan Lau[1,3]

[1]Department of Psychology, University of California Los Angeles, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563 USA; [2]Programs in Psychology and Biology, The Graduate Center of the City University of New York, 365 Fifth Ave., New York, NY 10016 USA; [3]Brain Research Institute, University of California Los Angeles, 695 Charles E. Young Dr. South, Los Angeles, CA 90095 USA

*Correspondence address. Department of Psychology, University of California Los Angeles, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563 USA, Tel: +1 323-596-1093, E-mail: meganakpeters@ucla.edu

Abstract

Response bias (or criterion) contamination is insidious in studies of consciousness: that observers *report* they do not see a stimulus may not mean they have absolutely no subjective experience; they may be giving such reports in relative terms in the context of other stimuli. Bias-free signal detection theoretic measures provide an excellent method for avoiding response bias confounds, and many researchers correctly adopt this approach. However, here we discuss how a fixation on avoiding criterion effects can also be misleading and detrimental to fruitful inquiry. In a recent paper, Balsdon and Azzopardi (Absolute and relative blindsight. *Consciousness and Cognition* 2015; **32**:79–91.) claimed that contamination by response bias led to flawed findings in a previous report of "relative blindsight". We argue that their criticisms are unfounded. They mistakenly assumed that others were trying (and failing) to apply their preferred methods to remove bias, when there was no such intention. They also dismissed meaningful findings because of their dependence on criterion, but such dismissal is problematic: many real effects necessarily depend on criterion. Unfortunately, these issues are technically tedious, and we discuss how they may have confused others to misapply psychophysical metrics and to draw questionable conclusions about the nature of TMS (transcranial magnetic stimulation)-induced blindsight. We conclude by discussing the conceptual importance of criterion effects in studies of conscious awareness: we need to treat them carefully, but not to avoid them without thinking.

Introduction

Blindsight is the phenomenon in which, following damage to the primary visual cortex, patients can display above-chance performance in discrimination or detection of visual stimuli despite their reported lack of conscious visual experiences (Weiskrantz, 1986, 1996; Cowey and Stoerig, 1991, 1995, 1997; Kentridge *et al.*, 1999, 2004; Cowey, 2010; Sahraie *et al.*, 2010; Overgaard, 2011; Ko and Lau, 2012; Kentridge, 2015). Studies of blindsight are central to our understanding of consciousness because unlike many other methods of manipulating perceptual awareness (e.g. masking), blindsight is not associated with completely abolished task performance capacity. This important dissociation allows us to avoid confounding awareness with just basic perceptual processing, and thereby to address the phenomenon of conscious awareness in a conceptually meaningful way (Lau, 2008; Lau and Rosenthal, 2011).

Because blindsight patients are rare, there have been attempts to recreate blindsight in healthy subjects (Kolb and Braun, 1995; Kunimoto *et al.*, 2001). Unfortunately, some of these efforts have turned out to be not replicable (Morgan *et al.*, 1997; Robichaud and Stelmach, 2003); others revealed the mathematical complexity of estimating the correspondence between confidence and accuracy (Galvin *et al.*, 2003; Evans and Azzopardi, 2007; Maniscalco and Lau, 2012). These efforts have also highlighted ongoing controversy regarding the relationship between metacognitive sensitivity (i.e. correspondence between confidence and accuracy) and conscious awareness (Charles *et al.*,
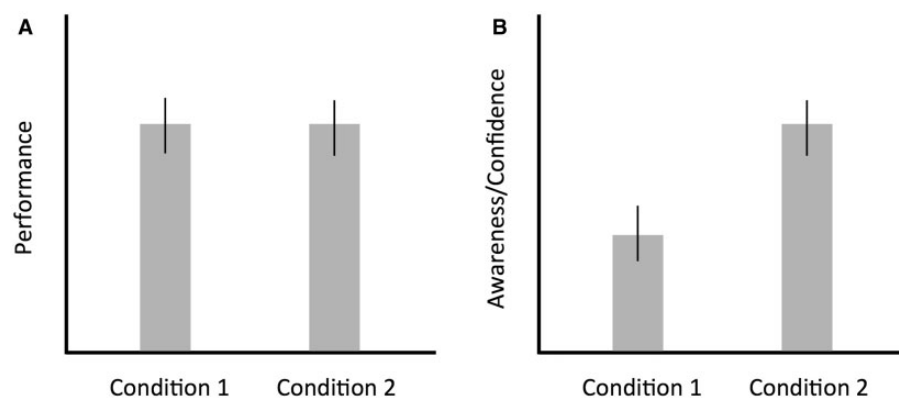
Figure 1. Illustration of a relative blindsight effect. For two conditions, Performance (indexed by percent correct, d', or some other measure of perceptual sensitivity/capacity) is equivalent, but Awareness or Confidence ratings differ.

2013; Fleming and Lau, 2014; Jachs *et al.*, 2015). Other groups have since attempted to experimentally induce milder forms of the phenomenon that are conceptually related but different. Below we refer to two such phenomena: "relative blindsight" (Lau and Passingham, 2006) and "TMS-induced blindsight" (Boyer *et al.*, 2005; Jolij and Lamme, 2005; Ro, 2010; de Graaf *et al.*, 2011).

Studies concerning both phenomena, however, have recently been criticized: it has been argued that both phenomena are susceptible to the contamination of response bias. This article evaluates and responds to these critiques. Importantly, although the last author (H.L.) was responsible for introducing the relative blindsight paradigm via metacontrast masking, his lab is currently no longer pursuing that line of research, mainly because the observed experimental effect was small (though replicable; Maniscalco and Lau, 2010; Maniscalco *et al.*, 2016) and there are likely better experimental paradigms to probe the same conceptual questions (e.g. Rounis *et al.*, 2010; Rahnev *et al.*, 2011; Koizumi *et al.*, 2015). As such, the main goal of this article is not to defend the existence and robustness of the phenomena in question. Rather, we hope to raise important issues concerning general attitudes toward response biases in perception research; ultimately, these issues concern far more than the phenomenon of blindsight itself.

Relative blindsight

Relative blindsight refers to the phenomenon that, for similar stimuli at identical objective task performance levels (e.g. accuracy in stimulus discrimination), observers can have different subjective levels (or frequency) of reported awareness in different conditions. That is, whereas actual clinical cases of blindsight involve the absolute abolishment of visual awareness while objective task performance is still above chance, relative blindsight involves a relative difference in subjective visual awareness levels while objective task performance is held constant (see Fig. 1).

The term "relative blindsight" was introduced by Lau and Passingham (2006) when they used metacontrast masking to demonstrate the phenomenon. They varied stimulus presentation timing parameters, specifically the stimulus onset asynchrony (SOA) between the visual target and the metacontrast mask, and identified two SOA conditions with identical objective performance levels in the task of discriminating between a square and a tilted square (a "diamond"). It was found that at the shorter of the two SOAs (in which the mask was presented

sooner after the target), subjects reported that they saw the target consciously less often than at the longer SOA, despite similar objective ("square or diamond?") task performance levels between the conditions.

The study (Lau and Passingham, 2006) has been criticized for a number of reasons. First Jannati and Dilollo (2012) argued that the "criterion content" may not be matched between the different SOA conditions. The term "criterion content" (Kahneman, 1968) here should not be confused with the concept of response criterion or bias. Rather, Jannati and Dilollo's (2012) point is that the objective discrimination task in Lau and Passingham (2006) may not have relied on the same visual information in the two matched-performance conditions. We do not address this point here because in a different study we replicated the phenomenon with stimuli for which criterion content is better matched (Maniscalco and Lau, 2010; Maniscalco *et al.*, 2016). Therefore, the possibility of differing criterion content is unlikely to be a critical issue.

Balsdon and Azzopardi (2015) also raised the issue of criterion content, together with other new criticisms. It was mentioned that stimulus timing was not carefully monitored in the original Lau and Passingham (2006) study. The last author (H.L.) duly acknowledges that this was the case. However, as Balsdon and Azzopardi (2015) pointed out, these are straightforward issues that can be fixed easily. They should not undermine the possibility that relative blindsight exists, but only concern the question of under what specific timing parameters the phenomenon can be obtained via metacontrast masking.

Balsdon and Azzopardi (2015) went on to raise what seems to be a more substantive issue, concerning the paradigm used by Lau and Passingham (2006): "Lau and Passingham reported comparing percent correct 2AFC scores for performance ('Was the target a square or a diamond?') and percent correct yes–no scores for awareness ('Did you see the target?').... [However,] the test that was used to assess performance was in fact not a 2AFC, ... but rather a yes-no question" (p. 81).

We do not dispute that a 2AFC task is different from a 2-choice discrimination task (or what Balsdon and Azzopardi (2015) call a yes-no, or YN, task, as explained in the next section). In the former, the same pair of stimuli is presented in every trial, and the subject must identify the spatial or temporal arrangement of the stimulus pair; in the latter, one out of two stimuli is presented in each trial and the subject must state which one of the two is presented. According to these standard definitions in psychophysics—which many researchers fail to adopt these days—2AFC includes a variety of tasks, e.g. a

2-interval forced choice detection task in which one has to identify whether a target is presented followed by a blank, or vice versa; a task to determine the spatial localization of a target to be presented on either the left or the right (the other location is occupied by "blank"); or a task in which the subject has to say whether a square is on the right and a diamond is on the left, or the other way round. These are all 2AFC tasks, but the task used in Lau and Passingham (2006) was decidedly not of the 2AFC variety because in each trial only one of two possible stimuli was presented—a square or a diamond. We do not dispute that it was a 2-choice discrimination and not a 2AFC. Yet the concern raised by Balsdon and Azzopardi (2015) is puzzling because at no point in Lau and Passingham (2006) did the authors mention a 2AFC task. Although Lau and Passingham (2006) did call their task a "forced-choice discrimination", nowhere did the authors claim they had used a 2AFC task. As we discuss later, these terminological confusions might more easily be avoided in the future if the term "forced-choice discrimination" were to be replaced by "2-choice discrimination".

We suspect Balsdon and Azzopardi's (2015) mis-reading of Lau and Passingham (2006) may also have stemmed from Balsdon and Azzopardi's (2015) belief that response bias is such a problematic contaminating factor (see also Evans and Azzopardi, 2007) that they expected Lau and Passingham (2006) to have used a 2AFC task as the objective discrimination task. It is true that 2AFC tasks are supposedly less susceptible to response bias (Macmillan and Creelman, 1990, 2004). However, as we will explain below, response bias is not really an issue in the objective task. So Lau and Passingham (2006) did not intend to use a 2AFC task at all; thus, there was no failure in implementing one. Yet it is worth noting that Balsdon and Azzopardi (2015) seem to be so concerned about response biases that they misread Lau and Passingham (2006), and so incorrectly reported Lau and Passingham's (2006) conclusions. How would response biases be such insidious, potentially "contaminating" factors?

A response bias problem?

Balsdon and Azzopardi (2015) called Lau and Passingham's (2006) objective task a yes-no (YN) task, which is technically correct. In this task an observer discriminates between a square and a diamond, and so one could rephrase the question as "Is the target a square or NOT?" in order to get "yes" versus "no" answers. (Incidentally, this method has also been called the "method of single stimuli" (Morgan et al., 2013), as the task is to discriminate a single stimulus.) However, we prefer to call it a "2-choice discrimination" task rather than a "yes-no" task because both stimulus possibilities are symmetrical in that they carry similar levels of physical intensity (and presumably overall neural activity too). This is quite different from a YN detection task in which the target-absent condition has lower stimulus energy. As we shall see below, this distinction may be important to avoid confusion. Nevertheless, let us examine whether using a 2-choice discrimination version of a YN task might possibly lead to response bias confounds.

In general, unlike in detection tasks, subjects tend not to show response biases in 2AFC tasks (e.g. Stanislaw and Todorov, 1999; but see also Cameron et al., 2002; Yeshurun et al., 2008; Morgan et al., 2013; and Acuna et al., 2015 for discussion on cases when 2AFC tasks can be susceptible to response biases, such as in the case of illusions). So, Balsdon and Azzopardi's (2015) challenge seems to be that Lau and Passingham's (2006) demonstration of relative blindsight may be trivially explained by response biases in the non-2AFC objective discrimination

task. But what difference could such response biases have made? One possible concern is that response bias can change the observed percent correct as a measure of performance, without altering underlying signal processing. That is, for two task conditions yielding identical underlying signal processing sensitivity, if one condition is associated with strong response bias but the other is not, percent correct ought to be lower in the condition associated with stronger response bias (Macmillan and Creelman, 2004). Therefore, if (a) Lau and Passingham's (2006) subjects showed response biases in discriminating square from diamond (i.e. they favored one over the other) *and* (b) this bias differed between SOA conditions, then task performance may not have been truly matched between the conditions as was reported. Yet this situation is highly implausible: it is unlikely that observers would show response biases for this kind of task, a 2-choice discrimination rather than detection, given the matched stimulus energies (Peters and Lau, 2015). And, even if subjects did show some bias, it seems unlikely that they would systematically show more or different bias in one SOA condition versus another. Therefore, Lau and Passingham (2006) did not consider it necessary to implement a 2AFC objective discrimination task to address this type of concern. In any case, the possibility of response bias contamination in the objective measure was subsequently addressed empirically: using a bias-free signal detection theoretic (SDT) measure of sensitivity (d') in the objective task, Maniscalco and colleagues (Maniscalco and Lau, 2010; Maniscalco et al., 2016) replicated Lau and Passingham's (2006) relative blindsight finding.

So, if anything, probably it is response bias in the *subjective* report that should be of concern here: is Lau and Passingham's (2006) finding of a difference in subjective confidence between matched-performance conditions potentially contaminated by response bias? To test this, Balsdon and Azzopardi (2015) designed an experiment which required subjects to discriminate between squares and diamonds under metacontrast masking, and to rate confidence afterwards, similar to Lau and Passingham's (2006) procedure. Balsdon and Azzopardi (2015) used four levels of difficulty, manipulated by varying the stimulus onset asynchrony (SOA) between target and mask for each subject. However, unlike Lau and Passingham (2006), they asked subjects to rate confidence on a scale of 1–4 (rather than "seen" or "guessed"). They subsequently defined the criterion for "seen" as a confidence rating of higher than 2, meaning a rating of 2 or less is defined as "unseen". At the easiest two levels of difficulty, they reported matched "square or diamond?" performance but differing percentages of "seen" versus "unseen", i.e. relative blindsight. Critically, they then examined the effect of shifting the criterion for "seen" versus "unseen" by grouping confidence ratings differently (e.g. [1] vs. [2, 3, 4] = liberal; [1, 2, 3] vs. [4] = conservative), and reported that at the most liberal criterion for "seen" (confidence >1), the relative blindsight effect vanished (p. 82). This finding is in line with a previous study, which showed that the magnitude of the relative blindsight effect is susceptible to shifting the criterion for "seen" (Jannati and Di Lollo, 2012).

However, these results are not surprising: shifting criteria can change the relative distribution of exemplars classified into two categories, and extreme criteria can obscure differences in group comparisons, even if significant differences do truly exist. For example, consider the (hypothetical) weight distributions between residents of the USA and Hong Kong (Fig. 2, top panel). If we take as the criterion for "overweight" a weight of 200 pounds (Fig. 2B), we will likely find many more overweight people (per capita) in the United States than in Hong Kong. But the percentage of people classified as overweight in each region
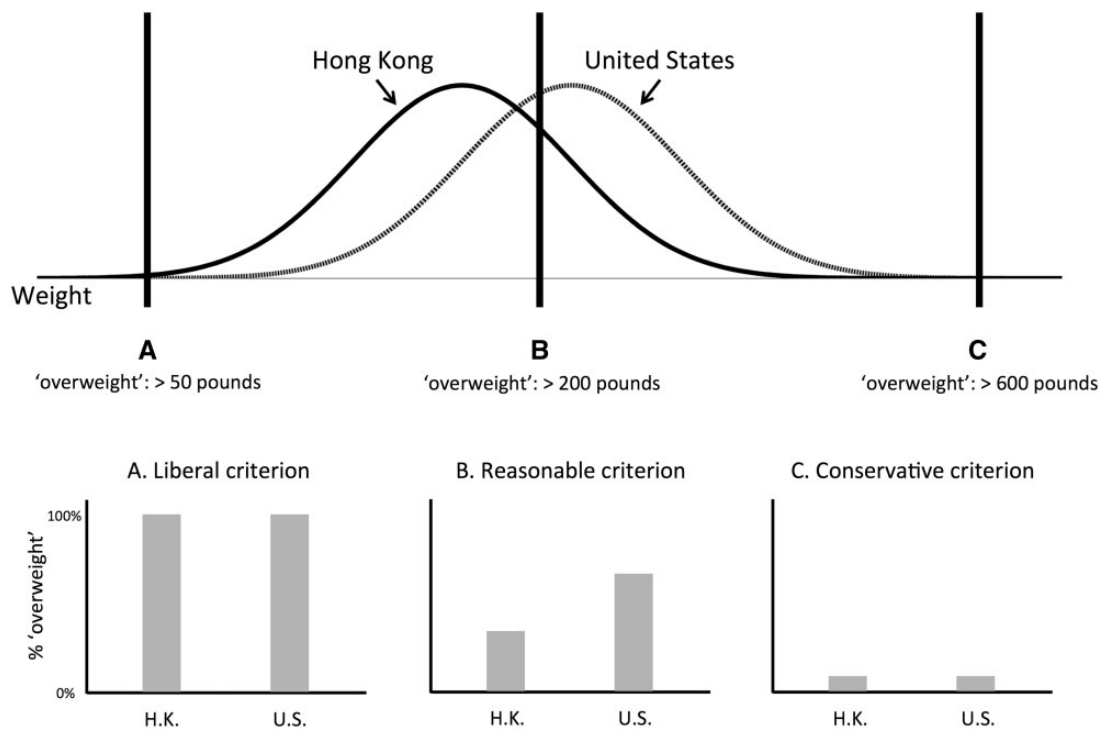
Figure 2. Illustration of hypothetical weight distributions of USA and Hong Kong citizens (top panel), and effects of different selections for a criterion to classify individuals as "overweight" (lower three panels). If we choose a criterion that is either too liberal (A) or too conservative (C), we will likely see no differences in the percentage of people classified as "overweight" between the USA and Hong Kong: nearly everybody or practically nobody will be classified as "overweight" in these two scenarios, respectively. But this does not mean the differences are not there; we must choose a reasonable criterion (B) in order to detect the differences. Note that Lau & Passingham (2006) encouraged subjects to use such a reasonable criterion for "awareness" when they reported a difference in "awareness" ratings across two conditions (e.g. Fig. 1B). However, their lack of finding any differences in objective performance across those two conditions (e.g. Fig. 1A) does not indicate that they chose an unreasonable criterion for performance: classifying trials as "correct" vs. "incorrect" does not depend on subjective criterion selection, and subjects did not display ceiling (100% correct) or floor (50% correct) performance. The relevant issue of whether response bias may play a role in performance measured here is discussed in the main text.

will also likely change somewhat if we shift this criterion, for example to 150 or 250 pounds. Considering an extreme case, if we use an overly conservative criterion, such as 600 pounds (Fig. 2C), we will likely find few if any differences between the two regions; the percentage of "overweight" people will approach zero in both regions under such an absurd definition. And likewise if we set an overly liberal criterion (as done by Balsdon and Azzopardi (2015) in "abolishing" the relative blindsight effect) of 50 pounds (Fig. 2A), we will likely find little difference in proportion of overweight people between the two regions, as presumably everybody will be classified as "overweight" (or nearly every stimulus classified as "seen"). Importantly, these changes in *observed* percentage of overweight individuals will occur despite no change in any underlying population differences in weight between the USA and Hong Kong. So, critically, we should not report as a result of an extreme criterion that there are no regional differences in proportion of individuals who are overweight, nor should we conclude that reporting such differences is incorrect because of the potential for criterion contamination. The difference is there, and it is up to the researchers to use an appropriate criterion level to reveal the difference (or to induce the subjects to do so, in a psychological experiment, especially in one in which subjects are asked to rate the subjective visibility of a stimulus). The same logic can apply to any situation in which one binarizes data.

So the claim that Lau and Passingham's (2006) observed difference in confidence between matched-performance conditions arises solely from response bias is misleading: of course finding the effect depends on bias/criterion selection, but this does not mean the effect is problematic or unreal. (In fact, under certain circumstances, subjects' criterion selections might themselves be informative.) Unfortunately this seems to be the main point of Balsdon and Azzopardi's (2015)'s challenge (though see Supplemental Material for other issues), and as such we think their worries are unfounded.

TMS-induced blindsight

The issue of response bias in the investigation of blindsight has also come up in other studies. For example, Boyer et al. (2005) demonstrated that transcranial magnetic stimulation (TMS) to the occipital pole (where V1/V2 lies) can induce blindsight-like behavior in normal observers. Subjects rated their awareness of a target ("yes" or "no"). If they were unaware of it, they were also required to discriminate either its orientation or color and rate their confidence in their decision. Even on trials in which they reported being unaware of the target, observers were able to discriminate its orientation and color above chance. Thus, the authors reported that TMS to V1 can produce temporary, blindsight-like behavior in normal observers. (See also Ro and Rafal, 2006, for further discussion.)

Lloyd *et al.* (2013) challenged this conclusion. They argued that TMS may simply raise the criterion for reporting a stimulus as "seen", and therefore induced not blindsight, but merely a response bias that primarily affects near-threshold stimuli. As such, Lloyd *et al.* (2013) argued that bias-free measures of detection sensitivity, such as the SDT measure d' (Green and Swets, 1966; Macmillan and Creelman, 2004), should be used to compare between YN awareness ratings and 2AFC discrimination tasks. So, as in Boyer *et al.* (2005), Lloyd *et al.* (2013) applied TMS to the occipital pole while subjects discriminated the targets and rated subjective awareness. Importantly, on some trials, the target was actually absent, which allowed them to use the bias-free SDT measure d' to assess detection sensitivity based on the awareness ratings—that is, to assess how well the awareness ratings tracked the presence or absence of the target (a Gabor patch). They also assessed discrimination sensitivity d' (left-tilted versus right-tilted orientation of the Gabor patch), and found that TMS impaired both the detection and discrimination sensitivities (d') in similar magnitudes. Based on this finding, they argued that TMS-induced blindsight is simply a case of near-threshold perception, in that there was no selective impairment of awareness that did not also impair objective information processing.

As we will argue below, Lloyd *et al.*'s (2013) interpretations raise important conceptual issues surrounding our understanding of blindsight and its relation to consciousness. First, however, we raise some methodological issues that may undermine their findings:

i. Lloyd *et al.* (2013) did not use a 2AFC measure of objective orientation discrimination performance, despite calling it a "forced-choice judgment (FC)" (p. 3) and later referring to it as a "2AFC" (p. 5). Because their task required observers to indicate orientation as a "left or right?" judgment, which is a 2-choice discrimination, it counts as a YN task in Balsdon and Azzopardi's (2015) terminology. As discussed elsewhere in this article, calling it a FC task may lead to confusions beyond merely terminological issues, so we advocate the use of "2-choice discrimination" rather than "forced-choice judgment" or "forced-choice discrimination" (as was unfortunately used by Lau and Passingham, 2006).

ii. Unlike Boyer *et al.* (2005) and other TMS-induced blindsight studies (e.g. Jolij and Lamme, 2005; Allen *et al.*, 2014), Lloyd *et al.* (2013) used near-threshold level visual stimuli for which objective discrimination performance was already low, even without any TMS. The addition of TMS to further reduce the low visibility of these near-threshold stimuli could have resulted in floor effects.

iii. These near-threshold stimuli may not have been strong enough to activate alternative visual pathways that may mediate some of the TMS-induced blindsight types measured in other studies (see also Ro *et al.*, 2004). The authors did not address this possibility.

iv. Lloyd *et al.* (2013) used a figure-eight TMS coil. Although a few studies (e.g. Kamitani *et al.*, 1999; Kammer *et al.*, 2005a, 2005b) have induced small visual scotomas using figure-eight coils at TMS intensities within the range of those used by Lloyd *et al.* (2013), the more focal stimulation induced with a figure-eight coil is not as effective at producing scotomas as compared to the larger circular coils (Kastner *et al.*, 1998), such as used by Boyer *et al.* (2005). Indeed, most TMS-induced blindsight studies have used circular coils to produce visual suppression (see also Jolij and Lamme, 2005; Christensen *et al.*, 2008; Allen *et al.*, 2014).

v. Unlike Boyer *et al.* (2005), who used TMS intensities that were 10% *above* the visual suppression threshold, Lloyd *et al.* (2013) used phosphene threshold TMS intensities in their study, which have been suggested to be substantially *lower* than the intensities required for producing scotomas with figure-eight coils (Kammer *et al.*, 2005a, 2005b).

Besides these issues, however, another concern with Lloyd *et al.*'s (2013) criticism of TMS-induced blindsight is conceptual: why would finding selective impairment in YN d' relative to 2AFC d' be a convincing demonstration of blindsight?

A psychophysical signature of consciousness impairment?

Lloyd and colleagues' (2013) emphasis on seeking a dissociation between YN and 2AFC sensitivity presumably rests on the observation of this effect in a blindsight patient. Azzopardi and Cowey (1997) compared YN d' and 2AFC d' in hemianopic blindsight patient G.Y. The authors used two tasks: one task was to indicate whether a target had been presented or not (YN detection), and another to indicate which of two sequentially-presented intervals contained a target ("first or second?" 2AFC). For these two corresponding tasks, 2AFC d' is known to be related to YN d' by a factor of $\sqrt{2}$ (Macmillan and Creelman, 2004); after this mathematical correction, control subjects demonstrated no differences between YN d' and 2AFC d'. On the other hand, patient G.Y. displayed smaller YN d' than 2AFC d' in his blind field (Azzopardi and Cowey, 1997). This approach has the advantage of comparing two d' values, which are supposedly bias-free (but see Yeshurun *et al.*, 2008). Yoshida and colleagues (Yoshida *et al.*, 2008) later replicated this finding in monkeys with unilateral V1 lesions, reporting a lower decision threshold in a 2AFC vs. a YN detection task, lending further support to use of this metric as an indicator of blindsight.

First we note that the mathematical relationship between YN d' and 2AFC d' does not always hold empirically (Macmillan and Creelman, 1990, 2004), so it is important to run control tasks and/or subjects. Importantly, that mathematical relationship holds only for true 2AFC tasks, so Lloyd *et al.*'s (2013) confusing 2AFC with 2-choice discrimination is problematic; this is a substantive technical issue that is beyond terminological choices. For their tasks we should not expect a fixed relationship, known *a priori*, between the sensitivity measures.

However, more importantly, why should we consider YN d' < 2AFC d' to be a *critical* signature of blindsight? Should it be considered a hallmark of impaired awareness or consciousness in general, to the point that we expect all demonstrations of unconscious perception to show this signature (c.f. Heeks and Azzopardi, 2015)?

Ultimately, for the phenomenon of blindsight to be meaningful in the context of consciousness studies, it cannot be identified purely via some idiosyncratic psychophysical signature. What then is the potential conceptual appeal of the metric? Perhaps it is because in YN detection, one has to distinguish between the presence and absence of an experience associated with a stimulus. One presumably performs the detection based on comparing the sense that something is presented to a lack of that subjective sensation. In 2AFC tasks, on the other hand, in both stimulus alternatives the "amount" of sensation is the same; one only has to determine the nature (temporal or spatial arrangement) of the stimuli. Perhaps in the 2AFC case, above-chance yet unconscious guessing *a la* blindsight is more likely, meaning that YN detection would be selectively impaired if

subjective awareness is abolished. Ko and Lau (2012) explored these issues and provided a simple computational framework (which relies on criterion effects) to account for YN d' < 2AFC d' as well as other psychophysical features of blindsight.

But if these conceptual considerations are important, it is not all instances of YN d' < 2AFC d' that we should be concerned with. It is the comparison between an estimate of YN *detection* d' and a corresponding 2AFC d' that is important here. As mentioned above, Azzopardi and colleagues prefer to call a 2-choice discrimination task a YN task, because it is not bias-free. In fact, by relying on this terminological equivalency, Balsdon and Azzopardi (2015) claimed they had abolished "absolute" blindsight (purportedly the difference between objective performance and subjective awareness; see Supplemental Material for further discussion) in a discrimination masking paradigm through demonstrating that "YN" (i.e. 2-choice discrimination) d' is equal to 2AFC d'. But it is unclear why one would expect to find YN d' < 2AFC d' under impairment of awareness if the YN task in question is *discrimination* rather than detection. This means the null findings of equal YN and 2AFC d' reported by Balsdon and Azzopardi (2015) may be trivial, because the YN task in question was discrimination, not detection.

The same logic holds for measures of 2AFC d': the important aspect is perhaps *not* that 2AFC tasks are bias-free, but rather that the two alternatives have similar levels of stimulus energy, so maintaining a criterion to discriminate between them seems relatively undemanding and the measured sensitivity for 2AFC tasks may be preserved in blindsight more so than for YN detection tasks. As discussed above, in this sense 2AFC tasks—detection or discrimination—are similar to 2-choice discrimination tasks, in which both stimulus alternatives also have similar levels of stimulus energy. Although many discrimination tasks can be framed as detections on the local scale (e.g. detection of cardinally-oriented lines in a "square or diamond?" task), the key difference between a detection and a 2-choice discrimination task is that the *overall* levels of stimulus energy are matched between the two stimulus alternatives. This makes it unlikely that subjects would use a "yes-no" kind of strategy (e.g. "yes I see cardinally-oriented lines so it is a square") in a 2-choice discrimination task. It seems that in developing the YN d' < 2AFC d' metric, the Azzopardi group (Azzopardi and Cowey, 1997, 1998; Evans and Azzopardi, 2007; Balsdon and Azzopardi, 2015; Heeks and Azzopardi, 2015) focused solely on the bias-free nature of 2AFC tasks compared to YN tasks, and did not consider the conceptual similarity between 2AFC and 2-choice discrimination, or the conceptual difference between detection and discrimination. We believe this is misguided, and reflects an over-emphasis on response bias.

Regardless of our take on the logic of Azzopardi and colleagues' method, it seems clear that Lloyd et al. (2013) were drawn to the YN d' < 2AFC d' metric because of the distinction between detection and discrimination, rather than because a 2AFC task is bias-free and a YN task is not. *Contra* Azzopardi and Cowey (1997, 1998), Lloyd et al. (2013) essentially (and perhaps unwittingly) compared d' in two YN tasks: YN *detection* versus 2-choice *discrimination* (rather than 2AFC). They probably did not see why a 2-choice discrimination task should be considered a YN task rather than 2AFC in this context.

So upon closer examination, we see that even for one of the few groups who have endeavored to implement the YN d' < 2AFC d' metric (Lloyd et al., 2013), the researchers did not actually agree with the logic and implementation of the original method. Other researchers investigating blindsight and its related phenomena in humans have simply not intended to rely

on the metric; thus, being criticized for failing to adopt it correctly is puzzling. Outside of blindsight studies, few authors refer to this metric at all. The moral seems to be that without a sound conceptual rationale, one can only go so far with a technical-sounding metric, even if coupled with the laudable intent of "response-bias avoidance."

Response biases and conscious perception

It is true that, historically, many have raised legitimate concerns regarding bias in statistical analysis or reporting of results. If, for example, we use "seen" reports from only one condition as a principled measure of conscious perception, we may be criticized on the grounds of criterion bias contamination, since every observer might have a different criterion for reporting "seen". So, many researchers try to control for the possibility of overly-conservative or liberal criterion-setting within an individual. These types of issues have been discussed at length in the study of semantic priming (Holender, 1992; Duscherer and Holender, 2005), unconscious processing (see Reingold and Merikle, 1990; Merikle and Reingold, 1998 for reviews), and indeed any task which can be analyzed with SDT metrics (Macmillan and Creelman, 1990, 2004).

Yet the study of *relative* blindsight is directed precisely at controlling for the potential confound of individual-level differences in criterion-setting: by comparing randomly interleaved conditions within the same session and same individual, any individual-level, global bias cannot account for performance differences across conditions. Instead, we can focus on condition-specific criterion effects, which should not be dismissed simply because they index biases or sheer difference in responding strategy. Importantly, criterion bias indexed by a SDT measure such as c or β cannot be definitively attributed to response-level effects (Witt et al., 2012, 2015). Certain perceptual phenomena—such as the sound-induced flash illusion (Shams et al., 2000; Shams, 2002), the ventriloquist effect (Howard and Templeton, 1966; Thurlow and Jack, 1973), the stream-bounce effect (Sekuler et al., 1997), and the Muller-Lyer illusion (Witt et al., 2012, 2015)—reveal themselves not only as differences in sensitivity across conditions, but as differences in measures of criterion as well. Thus, the criticism that TMS *merely* changed the criterion for subjective seeing does not necessarily undermine the phenomenon of TMS-induced blindsight (Ro et al., 2004; Boyer et al., 2005; Ro and Rafal, 2006). Changing the criterion may be an important perceptual phenomenon, both in cases of normal perception (e.g. Rahnev et al., 2011; Solovey et al., 2014) and in neurological cases of blindsight (Ko and Lau, 2012).

Traditionally, many researchers use SDT as a way to remove the potential "contamination" of criterion bias, but doing so should not be taken as evidence for technical or conceptual sophistication in all circumstances. If we are truly concerned with objective capacity differences among several conditions, we should certainly be cautious to avoid contamination of our comparison by bias. However, if we are concerned with the subjective rather than the objective aspects of perception, criterion bias may well be the very measure we should focus on. Therefore, depending on the research question, discarding bias can be just as thoughtless as throwing the baby out with the bathwater.

Acknowledgements

Supplementary data

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

References

Acuna DE, Berniker M, Fernandes HL, Kording KP. Using psychophysics to ask if the brain samples or maximizes. *J Vision* 2015;**15**:1–16.

Allen CPG, Sumner P, Chambers CD. The timing and neuroanatomy of conscious vision as revealed by TMS-induced blindsight. *J Cog Neurosci* 2014;**26**:1507–18.

Azzopardi P, Cowey A. Is blindsight like normal, near-threshold vision? *Proceedings of the National Academy of Sciences* 1997;**94**: 14190–4.

Azzopardi P, Cowey A. Blindsight and visual awareness. *Consciousness and Cognition* 1998;**7**:292–311.

Balsdon T, Azzopardi P. Absolute and relative blindsight. *Consciousness and Cognition* 2015;**32**:79–91.

Boyer JL, Harrison S, Ro T. Unconscious processing of orientation and color without primary visual cortex. *Proceedings of the National Academy of Sciences* 2005; **102**:16875–9.

Cameron EL, Tai JC, Carrasco M. Covert attention affects the psychometric function of contrast sensitivity. *Vision Res* 2002;**42**:949–67.

Charles L, Van Opstal F, Marti S, *et al.* Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage* 2013;**73**:80–94.

Christensen MS, Kristiansen L, Rowe JB, *et al.* Action-blindsight in healthy subjects after transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences* 2008;**105**: 1353–7.

Cowey A. Visual system: how does blindsight arise? *Current Biology* 2010;**20**:R702–R704.

Cowey A, Stoerig P. The neurobiology of blindsight. *Trends in Neurosciences* 1991;**14**:140–5.

Cowey A, Stoerig P. Blindsight in monkeys. *Nature* 1995;**373**: 247–9.

Cowey A, Stoerig P. Visual detection in monkeys with blindsight. *Neuropsychologia* 1997;**35**:929–39.

de Graaf Ta, Cornelsen S, Jacobs C, *et al.* TMS effects on subjective and objective measures of vision: Stimulation intensity and pre- versus post-stimulus masking. *Consciousness and Cognition* 2011;**20**:1244–55.

Duscherer K, Holender D. The role of decision biases in semantic priming effects. *Swiss Journal of Psychology* 2005;**64**:249–58.

Evans S, Azzopardi P. Evaluation of a "bias-free" measure of awareness. *Spatial Vision* 2007;**20**:61–77.

Fleming SM, Lau H. How to measure metacognition. *Frontiers in Human Neuroscience* 2014;**8**:1–9.

Galvin SJ, Podd JV, Drga V, *et al.* Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review* 2003;**10**: 843–76.

Green DM, Swets JA. *Signal Detection Theory and Psychophysics.* New York: John Wiley & Sons, Inc., 1966.

Heeks F, Azzopardi P. Thresholds for detection and awareness of masked facial stimuli. *Consciousness and Cognition* 2015;**32**: 68–78.

Holender D. Expectation effects, congruity effects, and the interpretation of response latency measurement. In: *Analytic Approaches to Human Cognition.* 1992, 351–75.

Howard IP, Templeton WB. *Human Spatial Orientation.* London: Wiley, 1966.

Jachs B, Blanco MJ, Grantham-hill S, *et al.* On the Independence of Visual Awareness and Metacognition: A Signal Detection Theoretic Analysis. *J Exp Psychol: Human Percept Perform* 2015;**41**:269–76.

Jannati A, Di Lollo V. Relative blindsight arises from a criterion confound in metacontrast masking: implications for theories of consciousness. *Consciousness and Cognition* 2012;**21**:307–14.

Jolij J, Lamme VAF. Repression of unconscious information by conscious processing: evidence from affective blindsight induced by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences* 2005;**102**:10747–51.

Kahneman D. Method, findings, and theory in studies of visual masking. *Psychol Bulletin* 1968;**70**: 404.

Kamitani Y, Kamitani Y, Shimojo S, *et al.* Manifestation of scotomas created by transcranial magnetic stimulation of human visual cortex. *Nature Neurosci* 1999;**2**:767–71.

Kammer T, Puls K, Erb M, *et al.* Transcranial magnetic stimulation in the visual system. II. Characterization of induced phosphenes and scotomas. *Exp Brain Res* 2005a;**160**:129–40.

Kammer T, Puls K, Strasburger H, *et al.* Transcranial magnetic stimulation in the visual system. I. The psychophysics of visual suppression. *Exp Brain Res* 2005b;**160**:118–28.

Kastner S, Demmer I, Ziemann U. Transient visual field defects induced by transcranial magnetic stimulation over human occipital pole. *Exp Brain Res* 1998;**118**:19–26.

Kentridge RW. What is it like to have type-2 blindsight? Drawing inferences from residual function in type-1 blindsight. *Consciousness and Cognition* 2015;**32**:41–4.

Kentridge RW, Heywood Ca, Weiskrantz L. Attention without awareness in blindsight. *Proceedings of the Royal Society B: Biological Sciences* 1999;**266**:1805–11.

Kentridge RW, Heywood Ca, Weiskrantz L. Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia* 2004;**42**:831–5.

Ko Y, Lau H A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 2012;**367**:1401–11.

Koizumi A, Maniscalco B, Lau H. Does perceptual confidence facilitate cognitive control? *Attention, Perception, & Psychophysics* 2015; doi:10.3758/s13414-015-0843-3

Kolb F, Braun J Blindsight in normal observers. *Nature* 1995;**377**:336–8.

Kunimoto C, Miller J, Pashler, H. Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition* 2001;**10**:294–340.

Lau H. Are we studying consciousness yet. *Frontiers of Consciousness: Chichele Lectures* 2008;**2008**: 245.

Lau H, Passingham RE Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences* 2006; **103**:18763–8.

Lau H, Rosenthal D. Empirical support for higher-order theories of conscious awareness. *TRENDS in Cogn Sci* 2011; **15**:365–73.

Lloyd DA, Abrahamyan A, Harris JA. Brain-stimulation induced blindsight: unconscious vision or response bias? *PLoS ONE* 2013;**8**:1–16.

Macmillan NA, Creelman CD. Response bias: characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychol Bulletin* 1990;**107**:401.

Macmillan NA, Creelman CD. *Detection Theory: A User's Guide*. Psychology Press, 2004.

Maniscalco B, Lau H. Comparing signal detection models of perceptual decision confidence. *Journal of Vision* 2010;**10**:213.

Maniscalco B, Lau H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition* 2012;**21**:422–30.

Maniscalco B, Peters MAK, Lau H. Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception & Psychophysics* 2016; doi:10. 3758/ s13414-016-1059-x.

Merikle PM, Reingold EM. On demonstrating unconscious perception: comment on Draine and Greenwald (1998). *J Exp Psychol: General* 1998;**127**:304–10.

Morgan MJ, Mason AJS, Solomon JA. Blindsight in normal subjects? *Nature* 1997;**379**:401–2.

Morgan MJ, Melmoth D, Solomon Ja. Linking hypotheses underlying Class A and Class B methods. *Visual Neurosci* 2013;**30**:197–206.

Overgaard M. Visual experience and blindsight: a methodological review. *Exp Brain Res* 2011;**209**:473–9.

Peters MAK, Lau H. Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *eLife* 2015; doi:10.7554/eLife.09651

Rahnev DA, Maniscalco B, Graves T, *et al*. Attention induces conservative subjective biases in visual perception. *Nature Neurosci* 2011;**14**:1513–5.

Reingold EM, Merikle PM. On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language* 1990;**5**:9–28.

Ro T. What can TMS tell us about visual awareness? *Cortex* 2010;**46**:110–3.

Ro T, Rafal R. Visual restoration in cortical blindness: insights from natural and TMS-induced blindsight. *Neuropsychol Rehab* 2006; **16**:377–96.

Ro T, Shelton D, Lee OL, *et al*. Extrageniculate mediation of unconscious vision in transcranial magnetic stimulation-induced blindsight. *Proceedings of the National Academy of Sciences* 2004;**101**:9933–5.

Robichaud L, Stelmach LB Inducing blindsight in normal observers. *Psychonomic Bulletin & Review* 2003;**10**:206–9.

Rounis E, Maniscalco B, Rothwell JC, *et al*. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cog Neurosci* 2010;**1**:165–75.

Sahraie A, Hibbard PB, Trevethan CT, *et al*. Consciousness of the first order in blindsight. *Proceedings of the National Academy of Sciences* 2010;**107**:21217–22.

Sekuler R, Sekuler aB, Lau R. Sound alters visual motion perception. *Nature* 1997. doi:10.1038/385308a0

Shams L. Integration in the brain- the subconscious alteration of visual perception by cross-modal integration. *Sci Con Rev* 2002;**October**:1–4.

Shams L, Kamitani Y, Shimojo S. Illusions: what you see is what you hear. *Nature* 2000; **408**:788.

Solovey G, Graney GG, Lau H. A decisional account of subjective inflation of visual perception at the periphery. *Attention, Perception, & Psychophysics* 2014;**77**:258–71.

Stanislaw H, Todorov N. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc* 1999;**31**:137–49.

Thurlow WR, Jack CE. Certain determinants of the "ventriloquism effect". *Perceptual and Motor Skills* 1973;**36**:1171–84.

Weiskrantz L. *Blindsight: A Case Study and Implications*. Oxford University Press, 1986.

Weiskrantz L. Blindsight revisited. *Current Opinion in Neurobiology* 1996;**6**:215–20.

Witt JK, Taylor E, Sugovic M, *et al*. Signal detection theory cannot distinguish perceptual and response-based biases: evidence from the Muller-Lyer illusion and application for action-specific effects. *Journal of Vision* 2012;**12**:899.

Witt JK, Taylor JET, Sugovic M, *et al*. Signal detection measures cannot distinguish perceptual biases from response biases. *Perception* 2015;**44**:289–300.

Yeshurun Y, Carrasco M, Maloney LT. Bias and sensitivity in two-interval forced choice procedures: tests of the difference model. *Vision Research* 2008;**48**:1837–51.

Yoshida M, Takaura K, Kato R, *et al*. Striate cortical lesions affect deliberate decision and control of saccade: implication for blindsight. *J Neurosci: Off J Society for Neurosci* 2008; **28**:10517–30.