8

THE OLD AND NEW CRITERION PROBLEMS

Matthias Michel

8.1 Introduction

To know what you see, I just have to ask. Assuming that you're not lying, I can interpret your report as an indicator of visual awareness. Your report is a reliable way of knowing about your perceptual state. Things aren't that easy once one enters the scientific lab. Researchers often present stimuli at the visibility threshold. They manipulate parameters such as the speed of presentation of stimulus, its contrast, whether or not the stimulus is closely followed by a masking stimulus, or whether the participant attends to the stimulus or not (Breitmeyer 2015; Kim and Blake 2005).

If you participate in one of those experiments, you'll often find yourself wondering whether you really saw the stimulus. In those cases, it is not clear what you should report. Should you answer "seen" or "not seen"? You need a criterion to determine what to answer in those borderline cases.

Here's one way to do it. You could create some sort of mental boundary, for instance, by saying to yourself, "Whenever the stimulus feels stronger than *that*, I'll say that I saw it, even though I'm not quite sure". *That* level of sensory strength, in other words, is your criterion. That boundary is somewhat arbitrary. Some people might be more *conservative* on what counts as "seen"—responding "seen" only when they're really sure they saw the stimulus. Others might settle for a liberal criterion—answering "seen" as soon as they felt like something was shown on the screen.

Enter the criterion problem. One can interpret negative reports, such as "not seen", as indicating either that the participant had no visual awareness of the stimulus, or as indicating that, although the participant did experience it, the strength of the stimulus fell below the subject's conservative criterion for

DOI: 10.4324/9781003409526-10

reporting “seen” (Reingold and Merikle 1988; 1990). The criterion problem is to decide between those two interpretations.

Sceptics argue that the criterion problem significantly prevents progress in consciousness research (Irvine 2012a, b; 2019), or that it should lead us to prefer procedures that do not rely on ‘subjective’ reports (Holender 1986; Phillips 2016; 2018; Reingold and Merikle 1990; Snodgrass et al. 2004). The alleged impossibility to distinguish changes in experience from changes in cognitive attitudes towards those experiences is also at the heart of the argumentative strategy for illusionism about consciousness (Dennett 1988; Frankish 2021, 60–61).

I hold that the sceptics should not be so pessimistic. We can experimentally distinguish between an absence of experience and an unwillingness to report having an experience. I present two solutions to this problem. The first solution relies on post-hoc data analyses. The second solution is to design bias-free tasks.

This leaves us with what I call the ‘new criterion problem’. Suppose that an experimental manipulation changes a participant’s tendency to report “seen”. How do you know whether that manipulation affected the participant’s experience of the stimulus, such that it now feels more visible, or simply their criterion for answering “seen”—making it more liberal?

Unlike the old criterion problem, there is no systematic solution to the new criterion problem. Still, I present two case studies indicating that it is possible to solve it on a case-by-case basis by being inventive. My conclusion is that criterion problems do not constitute good reasons for being sceptical of the value of subjective reports, and of the prospects of investigating what participants consciously experience.

8.2 Signal detection theory and the criterion problem

Suppose that you participate in an experiment in which your task is to detect a low-contrast stimulus. Signal Detection Theory (SDT) is a framework for understanding how you make this type of perceptual decision (Macmillan and Creelman 2005).

To see a stimulus, you need to be *perceptually sensitive* to that stimulus. You are perceptually sensitive to a stimulus when, in response to presentations of that stimulus, your perceptual system outputs *signals* that are on average stronger than the level of sensory activity present in the system when the stimulus is not present (also called the level of *noise*). The stronger these signals are compared to the noise alone, the higher your perceptual sensitivity.

But perceptual sensitivity alone is not enough. You also need a rule for selecting a response: How strong should sensory activity be before you answer that the stimulus was present? Answering that question amounts to setting a *response criterion*.

Criterion setting is flexible. If you're rewarded for answering that the stimulus was present when it was indeed present (hits), then you should set your criterion to maximize your hit rate. That rule should lead you to adopt a *liberal response criterion*. But if you are specifically instructed to avoid answering that the stimulus is present when it isn't, you should require a lot of evidence that the stimulus is present before answering that it is. You should adopt a *conservative* criterion for responding that the stimulus is present.

The beauty of the SDT framework is that perceptual sensitivity can be estimated independently of your response criterion (Macmillan and Creelman 2005). The measure of perceptual sensitivity is called *d'* ('d-prime'). The criterion measure is often noted *c*.

Researchers use this framework to understand how participants provide "seen"/"not seen" reports. Given the sensory evidence that there is a stimulus, the observer's response depends on the way in which her response criterion is set. A decision to report with a given response category is the conjoint product of sensory evidence and a response criterion.

For this reason, when a subject reports that she didn't see a stimulus, it could be *either* because she was not conscious of the stimulus, *or* because the strength of the sensory signal associated with the stimulus fell below her conservative criterion for reporting the stimulus as "seen", despite the fact that she consciously perceived it. In this latter case, scientists would be wrong to interpret her report as indicating that she was not conscious of the stimulus. That's the criterion problem.

Because these two interpretations of negative reports are available, what scientists routinely take as evidence for unconscious perception could simply result from subjects adopting a conservative bias and failing to report seeing stimuli that elicit only weak perceptual signals. For this reason, the criterion problem makes it difficult to rigorously demonstrate unconscious perception (Cheesman and Merikle 1984, 1986; Holender 1986; Irvine 2012a; Peters et al. 2017a; Phillips 2016; Merikle 1982, 1983, 1984).[1]

You'd be forgiven for thinking that this isn't a serious problem. The attitude we should adopt is to trust subjective reports unless we have good reason to doubt their accuracy. The participant answers "not seen", therefore, she didn't see the stimulus—end of story. But the problem is serious: We do have good reason to suspect that subjective reports are indeed overly conservative.

First, experimental evidence indicates that participants are not very good at adjusting their response criterion to multiple stimulus strengths (Gorea and Sagi 2000; 2002; Rahnev et al. 2011; Rahnev 2021; but see Denison et al. 2018). Instead, they probably set a single criterion across various stimulus strengths—or several criteria that 'attract' each other (Rahnev 2021). This leads to a conservative criterion for reporting weak stimuli because responses to weak stimuli are made with a criterion optimized for mid-range stimulus strength. Since most consciousness research experiments use various stimulus strengths within the same block of trials, this constitutes a good reason for holding that subjects have an overly conservative criterion for reporting weak stimuli.

Second, the 'Neyman-Pearson objective' (Green and Swets 1966). When performing tasks such as visual detection participants often have the (implicit) goal of minimizing the number of false alarms to a predetermined level—a goal known as the *Neyman-Pearson objective* (Curry et al. 1977; Green and Swets 1966). When sensitivity is low, the Neyman-Pearson objective leads to a conservative bias. To minimize the false-alarm rate, participants become more conservative as sensitivity decreases. In consciousness research, stimuli are presented at visibility threshold. So, sensitivity is typically low. Therefore, observers are more likely to adopt conservative biases.

The criterion problem is a credible threat for consciousness research. I do believe in unconscious perceptual effects (Michel 2022a). But whenever I see a strong unconscious perceptual effect—the kind that looks too good to be true—my first thought is, 'conservative criterion!' And I think it should be yours, too. The sceptics are probably right about this. But there's a difference between healthy scepticism and pessimism. I now explain how researchers solve the criterion problem.

8.3 Solving the criterion problem

There are two main ways to solve the criterion problem: Post-hoc statistical analyses, and bias-free tasks. I present both in turn.

8.3.1 A post hoc analysis-based solution

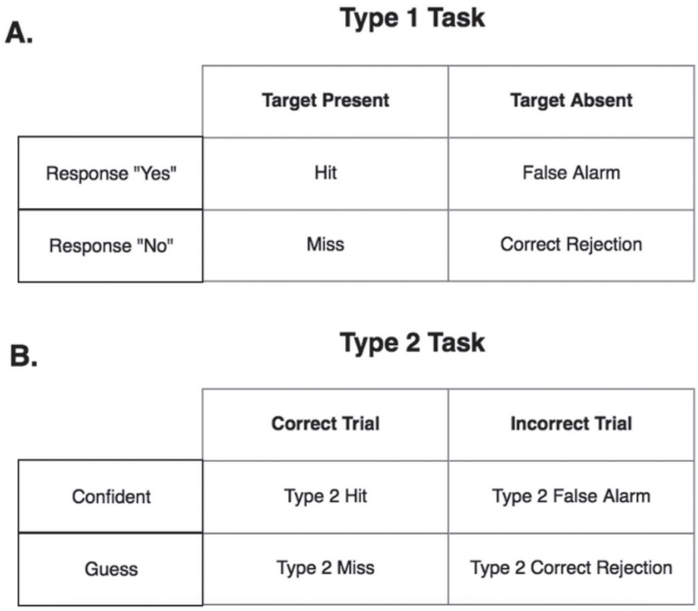
In consciousness research experiments, participants usually perform two tasks. In the *Type-1 task*, participants discriminate between different states of the external world. For instance, decide whether a stimulus is a square or a diamond. In the *Type-2 task*, subjects either assess their confidence in their performance on the Type-1 task with confidence ratings (Michel, 2022b), or determine whether or not they saw the stimulus by providing visibility ratings.

How well subjects perceive the stimuli can be determined by computing *d'*—an indicator of sensitivity. In a nutshell, this is done by comparing the rate of 'hits'—the subject correctly reports that the stimulus is a square and the rate of 'false alarms'—the subject reports that the stimulus is a square, but the stimulus was a diamond. Computing *d'* in this way allows us to have a measure of *perceptual sensitivity* that is independent of the subject's *response criterion*, namely, her tendency to report ‘square’ or ‘diamond’ irrespective of what she actually saw.

Based on the subjects’ responses on the Type-1 and Type-2 tasks, scientists can also determine their *metacognitive sensitivity* (Fleming and Lau 2014; Galvin et al. 2003; Maniscalco and Lau 2012; Nelson 1984; Michel 2022b). A good metacognitive observer knows that she correctly sees the stimuli when she does. She answers “seen”, or that she is confident in her responses, when her Type-1 performance indicates that she *did* see the stimuli, and that she did correctly perform the discrimination task.

A poor metacognitive observer doesn’t know whether she performs the task correctly or not. She tends to answer “not seen” even when her Type-1 performance indicates that she *does* see the stimuli, or she answers “random guess” even when her Type-1 performance indicates that she did *not* randomly guess. Poor metacognitive sensitivity also results from subjects being overly confident in mistaken Type-1 responses, or answering that they saw the stimuli when their Type-1 performance indicates that they did not.

Just as with Type-1 responses, Type-2 responses can be characterized in terms of ‘hits’, ‘false alarms’, ‘miss’, and ‘correct rejection’ (Figure 8.1) (Clarke et al. 1959; Galvin et al. 2003).



A. **Type 1 Task**

	Target Present	Target Absent
Response "Yes"	Hit	False Alarm
Response "No"	Miss	Correct Rejection

B. **Type 2 Task**

	Correct Trial	Incorrect Trial
Confident	Type 2 Hit	Type 2 False Alarm
Guess	Type 2 Miss	Type 2 Correct Rejection

FIGURE 8.1 (a) Classification of responses for Type 1 Detection. (b) Classification of responses for Type 2 Detection.

Source: Created by the author.

At this point, a difficulty is to take metacognitive *bias* into account. As Fleming and Lau (2014, 2) argue:

> Intuitively one can consider the extreme cases where subjects perform a task near threshold …, but rate every trial as low confidence, not because of a lack of ability to introspect, but because of an overly shy or humble personality. In such a case, the correspondence between confidence and accuracy is constrained by bias.

To see how researchers solve this challenge, let me introduce another concept, the Receiver Operating Characteristics (ROC) curve (Swets 1973).

Since *d'* is not contaminated by bias, the sensitivity of two subjects can be measured as having the same *d'* with different hit and false-alarm rates (Figure 8.2). An ROC curve is the set of possible hit/false-alarm rate pairs that an observer can produce with a given *d'*. And the observer's bias determines which of those possible pairs the subject actually realizes.

The shape of the ROC curve is independent from the observer's criterion setting. So, the area under the ROC curve (called the AUROC) provides a measure of the observer's sensitivity independently of her criterion setting.

In the same way, metacognitive sensitivity can be calculated on the basis of the subject's AUROC on the Type 2 task, called 'AUROC2' (Fleming and Lau 2014; Michel 2022b). If AUROC2 is 0.5, the subject has no metacognitive sensitivity, and if it is higher, the subject has some metacognitive sensitivity. Just as AUROC measures perceptual sensitivity free from Type-1 biases, AUROC2 measures metacognitive sensitivity free from Type-2 biases—it represents the observer's metacognitive sensitivity independently of her particular criterion setting on the Type-2 task.

Having a bias-free measure of metacognitive sensitivity is great. But what does all this have to do with *consciousness*? The heart of the current research program in consciousness science is the idea that conscious and unconscious information processing can be dissociated, at least to some extent (Baars 1986; LeDoux et al. 2020). Some tasks might require *conscious* perceptual information to be carried out, while others do not.

Providing *accurate introspective reports* about one's current conscious mental states is such a task: Introspective reports seem to overwhelmingly rely on conscious information.[2] One indication of this is that we are much better at providing accurate introspective reports about our conscious, compared to unconscious, mental states. However introspective reports are generated, the cognitive system that generates these reports must be much more sensitive to conscious perceptual information than to unconscious information. This means that information that is available to the introspective system is much more likely to be conscious than unconscious.

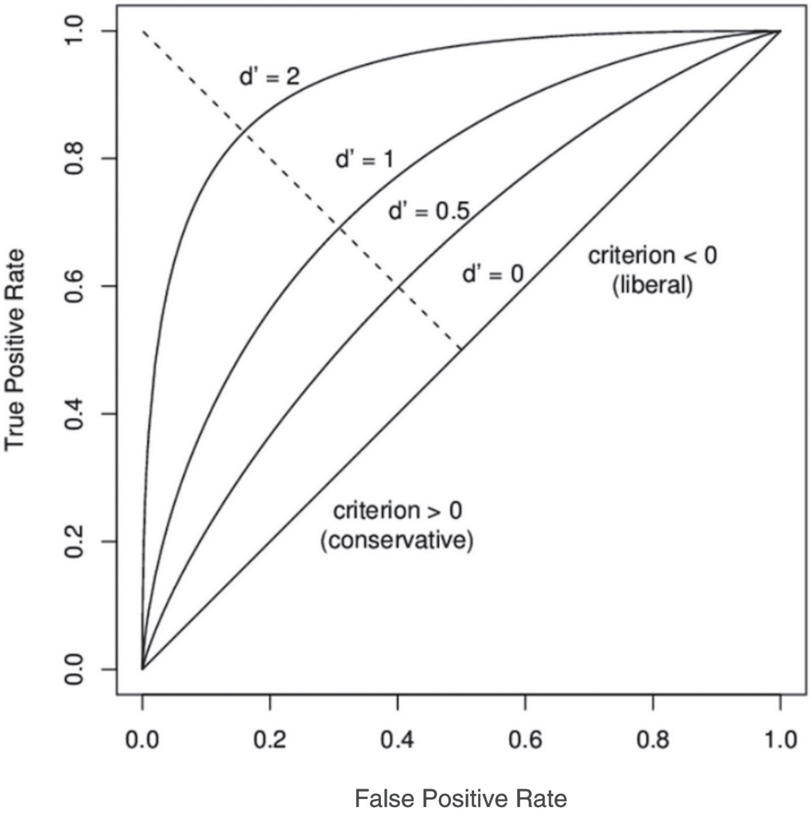


FIGURE 8.2 ROCs curves connect locations with constant *d'*. The diagonal is called the 'chance line': the hits and false alarm rates are equal. A conservative criterion decreases both the rates of hits and false alarms, and a liberal criterion has the opposite effect.

Source: Adapted from Macmillan and Creelman (2005).

High metacognitive sensitivity indicates that the subject is able to meaningfully modulate her introspective reports based on the visual information used in the Type-1 task. Given that the information used to successfully perform the Type-2 task is much more likely to be conscious than unconscious, this gives a good reason to hold that high metacognitive sensitivity indicates that the subject was conscious of the stimuli when the Type-1 task was performed.

In contrast, low metacognitive sensitivity could indicate that the subject didn't consciously see the stimuli during the Type-1 task.[3] One alternative interpretation would be that low metacognitive performance results from an overly conservative metacognitive response bias. But using the bias-free

AUROC2 to compute metacognitive sensitivity allows us to rule out this alternative explanation. The remaining explanation for the participant's inability to adapt her Type-2 responses to her Type-1 visual performance is that the information that she used to perform the Type-1 task was not conscious.

An alternative explanation for the participant's low metacognitive sensitivity could be that the subject was *phenomenally* conscious of the stimuli, without the subject having any *(meta)cognitive access* to that information (Block 1995; 2007). Phenomenal consciousness *overflows* the limits of metacognitive access.

I previously provided arguments against this view (Michel 2022b). Here, I simply want to distinguish this latter 'overflow' explanation from the alternative on which sensory signals simply fall below the conservative Type-2 response criterion. We should distinguish between the criterion problem and the overflow problem.

To introduce the distinction, take the following analogy. At home, I have a weight scale. That scale is biased: It always provides weight indications slightly under the true weight of the objects on the scale. If I were to blindly follow those indications, I might form the belief that some light objects actually have no weight at all! By calibrating the scale, and correcting for this bias, I would realize that this is not true.

But even with a perfectly calibrated scale, some objects might be so light that they do not activate the scale mechanisms at all. *This* limitation is not due to a bias. And solving the bias problem won't help. Some objects simply fall beyond the scope of what I can hope to measure with that scale. I need a different instrument to measure them.

In the same way, the overflow hypothesis postulates contents that fall beyond the kind of conscious contents that can be detected with procedures relying on subjective reports. On the other hand, the kind of conscious contents that give rise to the criterion problem *do* fall within the kind of conscious contents that *could* be detected by the procedure *if it were not biased*.

A solution to the criterion problem does not rule out the possibility that some phenomenally conscious contents are not cognitively accessed, and thus, go undetected by the procedure. The solution to the criterion problem described above allows scientists to rule out the possibility that a low metacognitive sensitivity due to Type-2 misses (i.e., a lot of "not seen" reports) is mainly the result of the subject's conservative Type-2 criterion. This solves the criterion problem, and that's the problem I focus on.

8.3.2 A task-based solution

Using AUROC2 as a bias-free indication of consciousness might seem like a statistical trick to sweep the problem under the rug. Scientists have found a way of computing metacognitive sensitivity independently of response biases. But the subjects who perform those tasks *are* still biased. If you're not satisfied with this, don't worry. There's another way to solve the problem: Transforming the Type-2 task into a 2-interval forced-choice task, or 2IFC task for short (Knotts et al. 2018; Peters and Lau 2015; Rajananda et al. 2018; see also Barthelme and Mamassian 2009; de Gardelle and Mamassian 2014).

In 2IFC paradigms, subjects perform a Type-1 discrimination task in two successive 'intervals', and make a Type 2 judgment comparing their introspective access to perceptual information in the two intervals.

Let me illustrate with a study by Peters and Lau (2015). Subjects saw a grating with either a left or a right tilt quickly followed by a mask, and then a second masked grating. After each stimulus presentation, or *interval*, subjects decided whether the stimulus was tilted left or right. Observers then indicated which of the two discrimination decisions they felt more confident in by betting on a specific discrimination decision (either first interval, or second interval).[4] Here's the crucial trick now: Unbeknownst to the subjects, some intervals did not contain a target, namely, Peters and Lau included target-absent intervals (Figure 8.3).

To understand how responses provided in 2IFC procedures can be used in consciousness detection procedures, think about the phenomenology that you would have if you unconsciously perceived the relevant feature. For you, it would feel just as if that feature had not been presented. Seeing the feature unconsciously feels like not seeing it.

If you are unable to discriminate between seeing the stimulus, and seeing no stimulus, then you did not see the stimulus consciously. The fact that you are unable to discriminate between the two indicates that, to you, seeing the stimulus just felt like seeing nothing.[5] As explained by Peters and Lau (2015, 3),

> if a certain above-chance discrimination seems introspectively no different from a random guess based on no stimulus at all (as reflected by betting behavior), we interpret the discrimination to be unconscious.

In the 'betting' version of the 2IFC paradigm, unconscious perception of the target occurs if the observer discriminates the target above chance while being unable to bet more often on the target-present interval than on the target-absent interval. In other words, observers unconsciously perceive the stimuli if they correctly discriminate the target, but bet randomly on the target-present interval versus the target-absent interval.

The same thing goes for versions of 2IFC paradigms relying on visibility ratings (e.g., Peters and Lau, 2015; Peters et al. 2017b). In this case, participants do not bet on the target-present interval versus the target-absent interval, but report in which interval the stimulus was 'more visible'. If participants perform the Type 1 task above chance, but cannot report that the target-present interval was more visible than the target-absent interval, this indicates

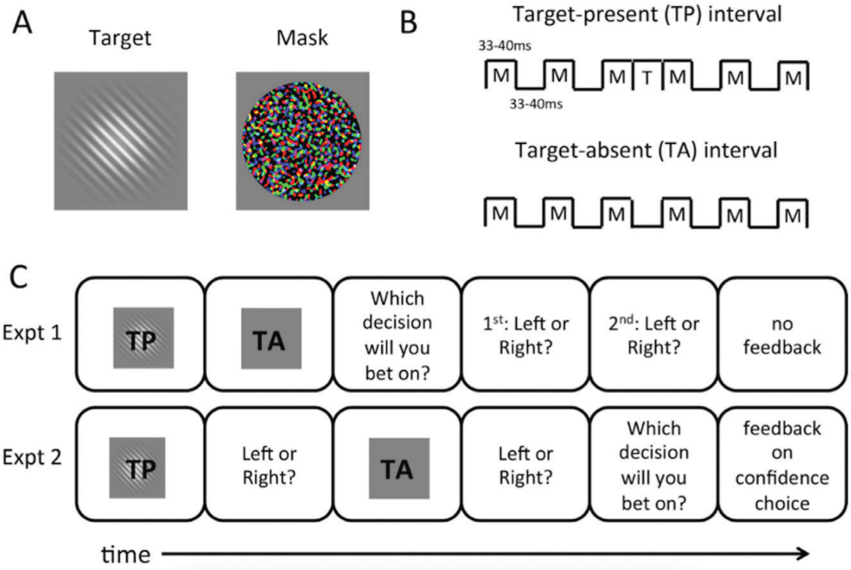


FIGURE 8.3 2IFC task. (a) Targets are oriented Gabor patches; masks are bandpass-noise filtered random RGB values. (b) Each trial consists of two intervals of discrimination. Some intervals contain a target (TP), while in others the target is replaced by a blank frame (TA). (c) Experimental tasks. In Experiment 1 subjects bet on which discrimination they feel more confident in before they indicate orientation discrimination choices (left or right tilt) for both intervals. The example presented is a trial in which TP is presented before TA; but in the experiment this order varied randomly. In Experiment 2, subjects bet on the more confident interval after the discriminations, and feedback is given.

Source: Peters and Lau (2015).

unconscious perception. The subjects did not judge the stimuli presented in the target-present interval to be more visible than no stimulus at all.[6]

2IFC paradigms are relatively free from biases (Green and Swets 1966; Macmillan and Creelman 2005). There's no reason to believe that conservative or liberal biases should lead subjects to systematically prefer responding that the stimulus is more visible in the first or second interval. For this reason, 2IFC paradigms 'discourage' biases (although, see Yeshurun et al. 2008). Since 2IFC paradigms are, if not completely bias-free, at least much less contaminated by biases, they allow researchers to solve the criterion problem.[7]

So, if the goal is to avoid biases and solve the criterion problem, researchers have at least two options at their disposal. For this reason, the criterion problem does not constitute a good reason for radical scepticism towards the prospects of a science of consciousness, since consciousness scientists *can* solve it.

8.4 The new criterion problem

Although researchers can ultimately solve the criterion problem, an additional problem with the criterion measure could turn out to be quite difficult to solve. I now present *the new criterion problem*. I then provide two case studies, based on experiments by Gallagher et al. (2019; 2021) and by Iemi and Busch (2018), suggesting that even this version of the problem can be solved.

8.4.1 What is the criterion all about?

Following a common interpretation of SDT, if the effect of an experimental manipulation is truly perceptual, it should change *d'*, and not the criterion measure, noted *c*. If an experimental manipulation affects *c* but not *d'*, experimenters infer that the manipulation did not affect perception, but instead changed the placement of the internal response criterion. In the same way, if an experimental manipulation changes Type 2 bias, but not Type 2 sensitivity, one can interpret this as indicating that the manipulation did not change the subject's conscious perception of the stimulus but, instead, her tendency to *respond* or *judge* that she saw the stimulus.

This common interpretation of SDT indicators is probably wrong. And the fact that it is wrong creates an underestimated difficulty. So, let me start by explaining why it is wrong.

Borrowing an example from Witt et al. (2015), we can use the Müller-Lyer illusion (Figure 8.4) to illustrate why differences in criterion setting (as measured by *c* in Signal Detection Theory) do not necessarily correspond to differences in response strategy, but also to perceptual differences.

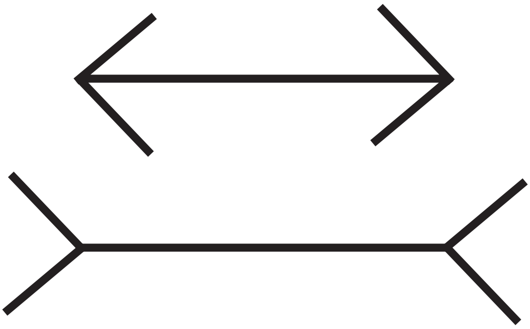


FIGURE 8.4 Müller-Lyer illusion. The lines are perceived as having different lengths but they have the same length. Lines with tails oriented inward are perceived as being shorter, while lines with tails oriented outward are perceived as being longer.

Source: Created by the author.

Imagine performing the following task: You see either short (5 cm) or long (7 cm) lines. Your goal is to discriminate between the two. Let's say that you are not biased: You set your response criterion optimally. When you perceive a line as being more than 6 cm, you answer that it is long. When you perceive it as being less than 6 cm, you answer that it is short (Figure 8.5a).

We can now see what happens if, for whatever reason, you prefer responding that the lines are shorter (Figure 8.5b). That is, you *perceive* lines between 6 and 7 cm as being long but, for whatever reason, *respond* that those lines are short. In this case, *d'* does not change—both hit and false-alarm rates increase at the same time. However, the *criterion* measure changes, compared to the initial situation.

At this point we introduce a slight change in the experiment. You set your response criterion exactly as in the first case—respond "long" whenever you *perceive* the line as being more than 6 cm. But now the lines have tails oriented inwards (Figure 8.4). When tails are oriented inwards, all the lines, short and long alike, look shorter. This leads you to answer "short" more often. And this is so even if your capacity to discriminate between short and long lines, as measured by *d'*, remains exactly the same as before.

In this case, a change in the *perception* of the lines manifests itself through the criterion measure *c*. As shown in Figure 8.5c, when the lines are perceived as being shorter (the signal distributions are switched to the left), *c* is identical to the case in which you do not *see* the lines as being shorter, but simply have a tendency to *respond* that the lines are short.

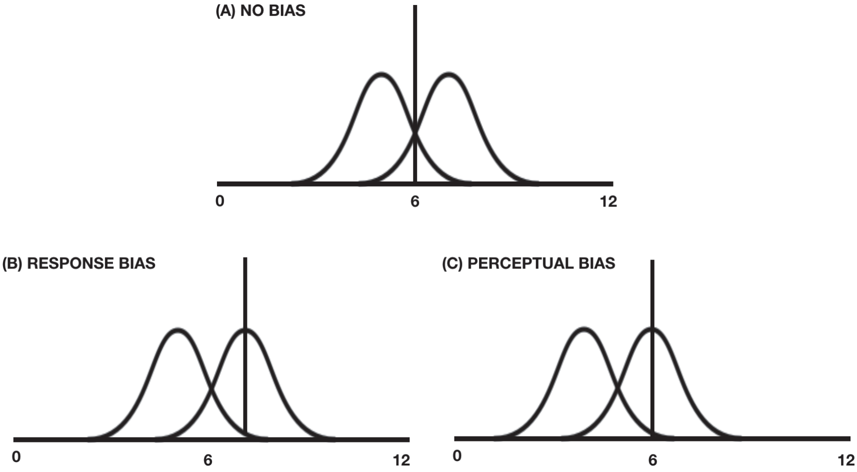


FIGURE 8.5 Hypothetical distributions of perceived line length for short and long lines. (a) Distributions without response bias. (b) Distributions with response bias. (c) Distributions with perceptual bias.

Source: Inspired from Witt et al. (2015).

If this is correct, perceptual effects sometimes manifest themselves through a measured change in *criterion*, and not *d'*. This goes against the view that changes in biases induced by experimental manipulations only reflect changes in response strategies, and not genuine changes in the way in which subjects *perceive* the stimuli.

Sánchez-Fuenzalida et al. (2022) confirmed the simulation work of Witt et al. (2015) by manipulating bias either with the Müller-Lyer illusion, or with unbalanced payoffs for answering "short" versus "long". These manipulations led to a criterion change when the participants reported their decisions. However, only the Müller-Lyer illusion led to a bias in a task where participants had to reproduce the perceived length of the lines—thus suggesting that the Müller-Lyer illusion elicits a criterion shift through a perceptual change.

In addition, several experiments have shown that well-known perceptual illusions, such as the sound-induced flash illusion (Shams et al. 2000; Shams 2002), the ventriloquist effect (Thurlow and Jack 1973), the stream-bounce effect (Sekuler et al. 1997; Meyerhoff & Scholl, 2018), and the Müller-Lyer illusion (Witt et al. 2015), not only lead to measured differences in sensitivity, but also to measured differences in criterion (see also, e.g., Polat and Sagi 2007). Linares et al. (2019) also found evidence that the criterion measure captures both perceptual and post-perceptual components even in simple discrimination tasks. Even better, using optogenetic manipulations of visual cortex activity in mice, Jin and Glickfeld (2019) induced changes in the criterion measure by directly manipulating sensory activity. Hence, one cannot interpret all measured changes in bias as response-strategy changes. Some changes in bias reflect perceptual effects (Peters et al. 2016; Witt et al. 2015).

Determining whether one should interpret variations in the criterion measure as reflecting *perceptual* or *response strategy* effects is what I call the new criterion problem.[8] In the following sections, I illustrate the new criterion problem with two case studies showing how the problem can be solved.

8.4.2 Solving the new criterion problem: Case study 1

A promising solution to the new criterion problem relies on the use of *confidence ratings*. Gallagher et al. (2019) asked subjects to report levels of confidence after performing a random dot motion task in which they had to discriminate between rightward or leftward motion. In one condition, Gallagher et al. induced a *perceptual* bias through sensory adaptation: after seeing dots moving in one direction, subjects tend to perceive dots as moving in the opposite direction. In the other condition, they induced a *response* bias by asking subjects to default to a given response direction, indicated by a cue, when hesitant about motion direction.

Gallagher et al. (2019) could reproduce similar psychometric curves by inducing either a perceptual bias, or a response bias. This is consistent with the new criterion problem: The discrimination task alone is not sufficient to differentiate between those two hypotheses (Morgan et al. 2012; Witt et al. 2015).

But things are different for confidence judgments. Gallagher et al. hypothesized that, while confidence judgments and the discrimination task both share the same perceptual evidence, those two types of judgments do not share the same response strategy parameters. As such, inducing a *perceptual* bias should equally influence discrimination *and* confidence judgments. But inducing a *response* bias should influence discrimination decisions without changing confidence measures.

Gallagher et al. confirmed this hypothesis. In the perceptual bias condition, following leftward sensory adaptation, subjects were biased towards answering that dots are moving to the right. They also needed less perceptual evidence for rightward movement, as determined by dot motion coherence, before answering that they were confident in their 'rightward movement' decisions. On the other hand, inducing a *response* bias only changed the perceptual decision bias without impacting the confidence measure. As Gallagher et al. (2019) conclude, dissociations between confidence judgments and discrimination performance could be used to distinguish between perceptual and response biases.

In subsequent experiments, Gallagher et al. (2021) used this method to solve the new criterion problem, as it applies to two perceptual effects: To biases induced by previous perceptual decisions, and to after-effects induced by implied motion.

In discrimination tasks, such as random dot-motion discrimination tasks, discrimination decisions are influenced by previous discrimination decisions, an effect known as the *choice history bias* (Fischer and Whitney 2014; Kanai and Verstraten 2005). We don't know whether this bias is perceptual or post-perceptual (but see Fritsche et al. 2017). Gallagher et al. (2021) used confidence ratings to distinguish between these two hypotheses.

They observed that discrimination decisions in a random dot motion task were influenced both by the previous trial (1-back effect) and by the trial-two trials prior (2-back effect). When subjects perceived leftward motion on the previous trial, they were biased towards answering "right" on the current trial. This bias also applied to confidence judgments: Subjects needed less perceptual evidence of a rightward motion before judging that they were confident. On the other hand, confidence ratings did not track the 2-back effect. These results are consistent with a *perceptual* effect of the previous trial on the current trial, akin to sensory adaptation, and a *post-perceptual* and a post-perceptual explanation of the 2-back effect.

The same method was applied to determine the nature of motion after-effects induced by *implied* movements in static images (Winawer et al. 2008; 2010). Gallagher et al. (2021) first obtained genuine motion after-effects: When subjects saw objects moving in one direction, their subsequent decisions were biased in the opposite direction. This after-effect also impacted the confidence measure. However, compared to actual movement, *implied* movements in static images influenced the discrimination decision, but did not impact the confidence measure. Following Gallagher et al., this result is consistent with a *post-perceptual* effect in the case of implied motion, compared to a *perceptual* effect in the case of real motion.[9]

Confidence and discrimination decisions rely on the same perceptual evidence. But they are not influenced by the same post-perceptual variables. It follows that manipulations that induce a bias in discrimination judgments without inducing a bias in the confidence judgments are more likely to stem from a *response* bias than a *perceptual* bias.

But this method is not a universal panacea. Experimental manipulations that induce similar biases on both Type-1 and Type-2 decisions remain ambiguous. They could be interpreted either as resulting from perceptual effects, or from response strategy effects influencing Type-1 and Type-2 criteria in the same way. In other words, this method is conclusive only if one observes a *dissociation* between Type-1 and Type-2 biases following an experimental manipulation. Still, this is better than nothing, and it suggests that the new criterion problem can be solved by using confidence ratings in a limited range of cases.

8.4.3 Solving the new criterion problem: Case study 2

Let me now turn to a second case study. When neurons are in a more excitable state before stimulus presentation, as indexed by the power of ongoing low-frequency oscillations (8-30 Hz)—weak α and β power reflecting high excitability—observers adopt more liberal biases in detection tasks, even if their perceptual accuracy remains the same (Ergenoglu et al. 2004; Benwell et al. 2017; Chaumon and Busch 2014; Limbach and Corballis 2016; Samaha et al. 2017). Researchers also observed that pre-stimulus neural excitability has an effect on detection tasks and not on discrimination tasks, thereby indicating that neural excitability affects bias, and not perceptual accuracy (Iemi et al. 2017).

Here, we face the new criterion problem: A change in bias could indicate either a *perceptual* change induced by weak pre-stimulus α and β power, or a change in *response strategy*. It could be that, when neurons are in a more excitable state, participants *feel* like they see stimuli more often. But it could also be that heightened excitability leads participants to *judge* that they see the target, and *answer* “yes, I saw the target”, more often.

Iemi and Busch (2018) designed an experiment to determine which interpretation is right, based on a 2IFC detection task, and a 2IFC discrimination task.

In the 2IFC *detection* task, a stimulus was presented in one of two intervals. Participants had to determine in which interval the stimulus was presented. In the 2IFC *discrimination* task, a target-stimulus (e.g., a left-tilted grating) was presented in one interval, and a non-target stimulus (e.g., a right-tilted grating) in the other interval. Participants had to report the interval in which the target-stimulus was presented.

Before we come to the predictions made by Iemi and Busch (2018), it is important to give some details on the psychological constructs posited by SDT to understand how participants perform 2IFC detection and discrimination tasks.

In 2IFC *detection* tasks, observers sample the internal response of a *single* feature detector in each interval, compare the internal responses of the detector in each interval, and then report that the stimulus was present in whichever interval yielded the strongest response. In 2IFC *discrimination* tasks, observers sample the internal responses of *two* feature detectors in each interval—one for the target, and one for the non-target stimulus. In each interval, the *relative* strengths of the responses for target versus non-target features are compared. The result of this comparison serves as evidence that the target was present in this interval (or not). Observers then compare the evidence in favour of the presence of the target between the two intervals and report that the target was present in the interval that had the strongest level of evidence for the presence of the target-stimulus.

With this, we can now turn to the predictions allowing Iemi and Busch (2018) to decide whether heightened excitability (i.e., weak pre-stimulus α and β power) leads to a liberal *perceptual* bias, or a liberal *response* bias.

A *perceptual* bias would increase the strength of the internal responses generated by the feature detectors. A *decision bias* would not affect the activity of the feature detectors, but only the observer's tendency to *judge* and *report* that the target was present in the interval with higher pre-stimulus excitability.

So, *if higher excitability leads to a perceptual bias* in the 2IFC *discrimination* task, the internal responses generated by the two detectors will increase, for *both* target and non-target features. Which means that the *relative* strength of the responses for target versus non-target features won't be affected. Observers have more evidence that the target-stimulus was present in the target-present interval, but they also have more evidence that the *non*-target-stimulus was present. As a result, higher excitability before the target-stimulus interval should not lead subjects to report that the target-stimulus was present in this interval more often.

In the 2IFC *detection task*, since there is only *one* feature detector, high excitability before the stimulus-present interval should lead subjects to report that the stimulus was present in this interval more often. As a result, the participants' hit rate should increase when pre-stimulus excitability is high in the target-present interval.

Now, compare this to the predictions *if higher excitability leads to a response bias*. In the discrimination task, the internal responses generated by the two detectors won't change. However, subjects will tend to *report* that the target-stimulus is present in whichever interval had higher pre-stimulus excitability. As a result, higher excitability before the target-stimulus interval should lead to a higher rate of hits. And the same thing goes for the detection task.

In sum, the hypothesis according to which higher pre-stimulus excitability induces a *perceptual bias* predicts a higher rate of hits in the *detection* task when pre-stimulus excitability is high before the stimulus-present interval, but not in the *discrimination* task when pre-stimulus excitability is high before the target-stimulus interval. On the other hand, if higher excitability induces a *response bias* one should observe a higher rate of hits in *both* the detection task, when pre-stimulus excitability is high before the stimulus-present interval, as well as in the discrimination task.

Iemi and Busch obtained results consistent with the hypothesis that higher excitability leads to a perceptual bias. Reduced α-and β-power (i.e., higher excitability) in the time window before the stimulus-present interval relative to the stimulus-absent interval led to a higher rate of hits in the detection task, but there was no statistically significant influence of reduced α-and β-power on the rate of hits in the discrimination task.[10] Iemi and Busch (2018, 21) conclude that "the current state of neuronal excitability—indexed by spontaneous α-and β-oscillations—biases the observer's subjective perceptual experience, by amplifying or attenuating sensory representations, rather than the decision strategy".

This study is a perfect example of the experimental ingenuity displayed in consciousness research. Of course, two case studies are not nearly enough to argue that similar strategies will be found in all cases in which one wants to decide whether a change in bias reflects a change in perception or in response strategy. But these case studies vindicate the view that the new criterion problem *can* be solved. In both cases, a solution to the new criterion problem could be found because, following some experimental manipulations, changes in perceptual and response criteria had different effects on the subjects' behavior. Several other experiments have successfully used this strategy to distinguish perceptual from post-perceptual biases (e.g., Fritsche et al. 2017; Linares et al. 2019; Sánchez-Fuenzalida et al. 2022). In the long run, if changes in perceptual and response biases have different effects, there will be someone inventive enough to develop an experiment teasing those effects apart, if not a strategy to systematically do so.

8.5 Conclusion

Irvine (2012b, 646) concluded her investigation of introspection-based methods in consciousness research in this way:

> The problems with report-based measures of consciousness have long been known and are again coming to light with the reintroduction of introspection-based measures. The recognition that reports are never free from bias must be made again, and the issues raised by the application of SDT to human perception must go through yet another round of reiteration. … These problems must be addressed in order to validate the use of introspective or subjective measures in consciousness science. However, given that this round of trying to make sense of the muddle surrounding subjective reports includes the same mistakes as the others that preceded it, abandoning the muddle altogether is looking more and more like the reasonable option to take.

While it is true that past mistakes have often been repeated, I believe that we can clear up the muddle. Neither the criterion problem nor the new criterion problem justify scepticism towards the prospects of consciousness research. Criteria effects are often confounding factors, but not always (Peters et al. 2016). And when they are, there is no reason to believe that consciousness scientists cannot find solutions to these problems. The criterion problem can be solved by using either post-hoc analyses, or 2IFC tasks. The new criterion problem cannot be solved in a *systematic* way at the moment. But several experimental manipulations allow for teasing apart the effects of perceptual and post-perceptual biases. Scientists can exploit those dissociations to determine whether biases are perceptual or post-perceptual in nature on a case-by-case basis.

Acknowledgments

I thank Steve Fleming, J.D. Knotts, Hakwan Lau, Brian Maniscalco, Jorge Morales, Kiyofumi Miyoshi, Brian Odegaard, and Ian Phillips for helpful comments and discussions. This work was funded by the Templeton World Charity Foundation, as part of the grant 'Analyzing and Merging Theories of Consciousness', at the Center for Mind, Brain, and Consciousness (New York University).

Notes

1. To the best of my knowledge, the first author to consider the criterion problem as a challenge for research on unconscious perception was Goldiamond (1958), following the rise of Signal Detection Theory and a series of studies by Blackwell (1952), who showed that threshold measurement in psychophysics is influenced by a variety of irrelevant factors. The problem was also mentioned early on by Eriksen (1960).
2. See Michel (2022b) for a defense of the view that explicit metacognitive mechanisms are 'consciousness selective'—they only take conscious sensory evidence as inputs.

3 This interpretation is available only if this low metacognitive performance is mainly due to a high rate of Type-2 misses. A high rate of Type-2 false alarms would also decrease the level of metacognitive sensitivity, but would rather indicate that the subject hallucinated non-existent stimuli, or that the subject simply performed the Type-2 task randomly.

4 Peters and Lau (2015) also included a control experiment in which subjects had to provide reports akin to visibility ratings instead of confidence ratings. In this variant of the experiment, participants had to indicate which of the two intervals was more visible. The results are the same.

5 This inference is warranted only if one assumes that subjects do not hallucinate targets in the target-absent intervals.

6 Macmillan (1986, 38) suggested something very similar as the best operational definition of the term ‘subliminal’: “[A] subliminal stimulus is one that leads to the detect state just as often as does a null stimulus.” According to this view, a stimulus is not consciously perceived if it is indistinguishable from no stimulus at all (the null stimulus).

7 I have previously criticized the paradigm used by Peters and Lau (2015) as being overly conservative about what counts as unconscious perception (Michel 2022a). Still, it remains true that the 2IFC paradigm solves the criterion problem—even if in that particular instance it was unsuccessful at identifying unconscious perception.

8 While the problem has been underestimated in the study of consciousness, it is far from being a *new* problem in the study of *perception* more broadly. Fechner already mentioned the influence of the subjects’ decision strategies on psychophysical judgments (Fechner 1860/1966), and the issue was discussed by Sekuler and Erlebacher (1971) in the case of judgments of subjective equality. More recently, both Morgan et al. (2012) and Witt et al. (2015) argued that 2-alternative forced choice tasks cannot distinguish perceptual from response criteria effects (see also García-Pérez and Alcalá-Quintana 2013; Gold and Ding 2013; Sánchez-Fuenzalida et al. 2022). We also find the problem arising in many subdomains of perceptual psychology, for instance, in studies on time perception (Yarrow et al. 2011), on biases induced by previous perceptual decisions on currently perceived stimuli (e.g., St-John-Saaltink et al. 2016; Fritsche et al. 2017), on the effects of attention on judgments of length (Milner et al. 1992), or other appearance judgments such as perceived contrast (Carrasco et al. 2004; 2008; Prinzmetal et al. 2008).

9 See Mather and Sharman (2015) for a similar result with a different method. See also Witthoft et. al (2018) for a study trying to dissociate perceptual and decision biases in a different kind of after-effects by using response time measures.

10 Because a null result does not indicate that the null hypothesis (i.e., no effect of excitability on accuracy) is true, Iemi and Busch (2018) also computed Bayesian Factors to determine the plausibility of the null hypothesis in the case of the discrimination task. They report that “The [Bayesian Factors] analysis revealed that the proportion of data points in favour of a null effect on 2IFC discrimination accuracy by far outnumbered the proportion of data points in favour of an effect” (Iemi and Busch 2018, 15), which indicates that the null hypothesis was probably true: There was no effect of reduced α- and β-power on accuracy in the discrimination task.

References

Baars, Bernard J. 1986. “What Is a Theory of Consciousness a Theory of?: The Search for Criterial Constraints on Theory.” *Imagination, Cognition and Personality* 6 (1): 3–23. https://doi.org/10.2190/WJER-XABV-QM4W-KD6V

Baars, Bernard J. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Barthelme, Simon, and Pascal Mamassian. 2009. “Evaluation of Objective Uncertainty in the Visual System.” *PLOS Computational Biology* 5 (9): 1–8. https://doi.org/10.1371/journal.pcbi.1000504

Benwell, Christopher S. Y., Chiara F. Tagliabue, Domenica Veniero, Roberto Cecere, Silvia Savazzi, and Gregor Thut. 2017. “Prestimulus EEG Power Predicts Conscious Awareness But Not Objective Visual Performance.” *ENeuro* 4 (6): 1–17. https://doi.org/10.1523/ENEURO.0182-17.2017

Breitmeyer, Bruno G. 2015. “Psychophysical ‘Blinding’ Methods Reveal a Functional Hierarchy of Unconscious Visual Processing.” *Consciousness and Cognition* 35: 234–250. https://doi.org/10.1016/j.concog.2015.01.012

Blackwell, H. Richard. 1952. “Studies of Psychophysical Methods for Measuring Visual Thresholds.” *Journal of the Optical Society of America* 42 (9): 606–616. https://doi.org/10.1364/josa.42.000606

Block, Ned. 1995. “On a Confusion about a Function of Consciousness.” *Behavioral and Brain Sciences* 18 (2): 227–247. https://doi:10.1017/S0140525X00038188

Block, Ned. 2007. “Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience.” *Behavioral and Brain Sciences* 30 (5–6): 481–499. https://doi.org/10.1017/S0140525X07002786

Carrasco, Marisa, Stuart Fuller, and Sam Ling. 2008. “Transient Attention Does Increase Perceived Contrast of Suprathreshold Stimuli: A Reply to Prinzmetal, Long, and Leonhardt.” *Perception and Psychophysics* 70 (7): 1151–1164. https://doi.org/10.3758/pp.70.7.1151

Carrasco, Marisa, Sam Ling, and Sarah Read. 2004. “Attention Alters Appearance.” *Nature & Neuroscience* 7 (3): 308–313. https://doi.org/10.1038/nn1194

Chaumon, Maximilien, and Niko A. Busch. 2014. “Prestimulus Neural Oscillations Inhibit Visual Perception Via Modulation of Response Gain.” *Journal of Cognitive Neuroscience* 26 (11): 2514–2529. https://doi.org/10.1162/jocn_a_00653

Cheesman, Jim, and Philip M. Merikle. 1984. “Priming With and Without Awareness.” *Perception & Psychophysics* 36 (4): 387–395. https://doi.org/10.3758/BF03202793

Cheesman, Jim, and Philip M. Merikle. 1986. “Distinguishing Conscious from Unconscious Perceptual Processes.” *Canadian Journal of Psychology* 40 (4): 343–367. https://doi.org/10.1037/h0080103

Clarke, Frank R., Theodore G. Birdsall, and Wilson P. Tanner. 1959. “Two Types of ROC Curves and Definitions of Parameters.” *The Journal of the Acoustical Society of America* 31 (5): 629–630. https://doi.org/10.1121/1.1907764

Curry, Renwick E., David C. Nagel, and Eliezer G. Gai. 1977. “Decision Behavior with Changing Signal Strength.” *Journal of Mathematical Psychology* 15 (1): 46–69. https://doi.org/10.1016/0022-2496(77)90040-2

de Gardelle, Vincent, and Pascal Mamassian. 2014. “Does Confidence Use a Common Currency Across Two Visual Tasks?” *Psychological Science* 25 (6): 1286–1288. https://doi.org/10.1177/0956797614528956

Denison, Rachel, William T. Adler, Marisa Carrasco, and Wei Ji Ma. 2018. "Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence." *Proceedings of the National Academy of Sciences of the United States of America* 115 (43): 11090–11095. https://doi.org/10.1073/pnas.1717720115

Dennett, Daniel C. 1988. "Quining Qualia." In *Consciousness in Contemporary Science*, edited by Anthony J. Marcel, and Eduard Bisiach, 42–77. Oxford: Oxford University Press.

Ergenoglu, Tolgay, Tamer Demiralp, Zubeyir Bayraktaroglu, Mehmet Ergen, Huseyin Beydagi, and Yagis Uresin. 2004. "Alpha Rhythm of the EEG Modulates Visual Detection Performance in Humans." *Cognitive Brain Research* 20 (3): 376–383. https://doi.org/10.1016/j.cogbrainres.2004.03.009

Eriksen, Charles W. 1960. "Discrimination and Learning without Awareness: A Methodological Survey and Evaluation." *Psychological Review* 67 (5): 279–300. https://doi.org/10.1037/h0041622

Fechner, Gustav T. 1860/1966. *Elements of Psychophysics.* New York: Holt.

Fischer, Jason, and David Whitney. 2014. "Serial Dependence in Visual Perception." *Nature Neuroscience* 17 (5): 738–743. https://doi.org/10.1038/nn.3689

Fleming, Stephen M., and Hakwan C. Lau. 2014. "How to Measure Metacognition." *Frontiers in Human Neuroscience* 8 (July): 443. https://doi.org/10.3389/fnhum.2014.00443

Frankish, Keith. 2021. "Panpsychism and the Depsychologization of Consciousness." *Aristotelian Society Supplementary Volume* 95 (1): 51–70. https://doi.org/10.1093/arisup/akab012

Fritsche, Matthias, Pim Mostert, and Floris P. de Lange. 2017. "Opposite Effects of Recent History on Perception and Decision." *Current Biology* 27 (4): 590–595. https://doi.org/10.1016/j.cub.2017.01.006

Galvin, Susan J., John V. Podd, Vit Drga, and John Whitmore. 2003. "Type 2 Tasks in the Theory of Signal Detectability: Discrimination Between Correct and Incorrect Decisions." *Psychonomic Bulletin & Review* 10 (4): 843–876. https://doi.org/10.3758/bf03196546

García-Pérez, Miguel. A., and Rocío Alcalá-Quintana. 2013. "Shifts of the Psychometric Function: Distinguishing Bias from Perceptual Effects." *Quarterly Journal of Experimental Psychology* 66 (2): 319–337. https://doi.org/10.1080/17470218.2012.708761

Gallagher, Regan M., Thomas Suddendorf, and Derek H. Arnold. 2019. "Confidence as a Diagnostic Tool for Perceptual Aftereffects." *Scientific Reports* 9 (7124): 1–12. https://doi.org/10.1038/s41598-019-43170-1

Gallagher, Regan M., Thomas Suddendorf, and Derek H. Arnold. 2021. "The Implied Motion Aftereffect Changes Decisions, but Not Confidence." *Attention, Perception, and Psychophysics* 83 (8): 3047–3055. https://doi.org/10.3758/s13414-021-02331-z

Gold, Joshua I., and Long Ding. 2013. "How Mechanisms of Perceptual Decision-Making Affect the Psychometric Function." *Progress in Neurobiology* 103: 98–114. https://doi.org/doi:10.1016/j.pneurobio.2012.05.008

Goldiamond, Israel. 1958. "Indicators Of Perception: I. Subliminal Perception, Subception, Unconscious Perception: An Analysis in Terms of Psychophysical Indicator Methodology." *Psychological Bulletin* 55 (6): 373–411. https://doi.org/10.1037/h0046992

Gorea, Andrei, and Dov Sagi. 2000. "Failure to handle more than one internal representation in visual detection tasks." *Proceedings of the National Academy of Sciences of the United States of America* 97 (22): 12380–12384. https://doi.org/10.1073/pnas.97.22.12380

Gorea, Andrei, and Dov Sagi. 2002. "Natural extinction: A Criterion Shift Phenomenon." *Visual Cognition* 9 (8): 913–936. https://doi.org/10.1080/13506280143000638

Green, David M., and John A. Swets. 1966. *Signal Detection Theory and Psychophysics.* New York: Wiley.

Holender, Daniel. 1986. "Semantic Activation Without Conscious Identification in Dichotic Listening, Parafoveal Vision, and Visual Masking: A Survey and Appraisal." *Behavioral and Brain Sciences* 9: 1–23. https://doi.org/10.1017/S0140525X00021269

Iemi, Luca, and Nico A. Busch. 2018. "Moment-To-Moment Fluctuations In Neuronal Excitability Bias Subjective Perception Rather Than Decision-Making." *ENeuro* 5 (3): ENEURO.0430-17.2018. https://doi.org/10.1523/ENEURO.0430-17.2018

Iemi, Luca, Maximilien Chaumon, Sébastien Crouzet, and Niko A. Busch. 2017. "Spontaneous Neural Oscillations Bias Perception by Modulating Baseline Excitability." *Journal of Neuroscience* 37 (4): 807–819. https://doi.org/10.1523/JNEUROSCI.1432-16.2016

Irvine, Elizabeth. 2012a. *Consciousness as a Scientific Concept*. Springer.

Irvine, Elizabeth. 2012b. "Old Problems with New Measures in the Science of Consciousness." *British Journal for the Philosophy of Science* 63 (3): 627–648. https://doi.org/10.2307/23253415

Irvine, Elizabeth. 2019. "Developing Dark Pessimism Towards the Justificatory Role of Introspective Reports." *Erkenntnis* 86: 1319–1344. https://doi.org/10.1007/s10670-019-00156-9

Jin, Miaomiao, and Lindsey L. Glickfeld. 2019. "Contribution of Sensory Encoding to Measured Bias." *Journal of Neuroscience* 39 (26): 5115–5127. https://doi.org/10.1523/JNEUROSCI.0076-19.2019

Kanai, Ryota, and Frans A. J. Verstraten. 2005. "Perceptual Manifestations of Fast Neural Plasticity: Motion Priming, Rapid Motion Aftereffect and Perceptual Sensitization." *Vision research* 45 (25–26): 3109–3116. https://doi.org/10.1016/j.visres.2005.05.014

Kim, Chay-Youn, and Randolph Blake. 2005. "Psychophysical Magic: Rendering the Visible 'Invisible'." *Trends in Cognitive Sciences* 9 (8): 381–388. https://doi.org/10.1016/j.tics.2005.06.012

Knotts, Jeffrey D., Hakwan Lau, and Megan A. K. Peters. 2018. "Continuous Flash Suppression and Monocular Pattern Masking Impact Subjective Awareness Similarly." *Attention, Perception, & Psychophysics* 80 (8): 1974–1987. https://doi.org/10.3758/s13414-018-1578-8

LeDoux, Joseph E., Matthias Michel, and Hakwan Lau. 2020. "A Little History Goes a Long Way Toward Understanding Why We Study Consciousness the Way We Do Today." *PNAS* 117: 6976–6984. https://doi.org/10.1073/pnas.1921623117

Limbach, Katharina, and Paul M. Corballis. 2016. "Prestimulus Alpha Power Influences Response Criterion in a Detection Task." *Psychophysiology* 53 (8): 1154–1164. https://doi.org/10.1111/psyp.12666

Linares, Daniel, David Aguilar-Lleyda, and Joan López-Moliner. 2019. “Decoupling Sensory from Decisional Choice Biases in Perceptual Decision Making.” *ELife* 8 (e43994): 1–22. https://doi.org/10.7554/eLife.43994

Macmillan, Neil A. 1986. “The Psychophysics of Subliminal Perception.” *Behavioral and Brain Sciences* 9: 38–39. https://doi.org/10.1017/S0140525X00021427

Macmillan, Neil A., and Douglas C. Creelman. 2005. *Detection Theory: A User’s Guide*. New York: Psychology Press.

Maniscalco, Brian, and Hakwan Lau. 2012. “A Signal Detection Theoretic Approach for Estimating Metacognitive Sensitivity from Confidence Ratings. *Consciousness and Cognition* 21 (1): 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Mather, George, and Rebecca J. Sharman. 2015. “Decision-Level Adaptation In Motion Perception.” *Royal Society Open Science* 2 (12): 150418. https://doi.org/10.1098/rsos.150418

Merikle, Philip M. 1982. “Unconscious Perception Revisited.” *Perception & Psychophysics* 31 (3): 298–301. https://doi.org/10.3758/BF03202538

Merikle, Philip M. 1983. “Subliminal Perception Reaffirmed.” *Canadian Journal of Psychology* 37 (2): 324–326. https://doi.org/10.1037/h0080891

Merikle, Philip M. 1984. “Toward a Definition of Awareness.” *Bulletin of the Psychonomic Society* 22 (5): 449–450. https://doi.org/10.3758/BF03333874

Meyerhoff, Hauke S., and Brian J. Scholl. 2018. “Auditory-induced bouncing is a perceptual (rather than a cognitive) phenomenon: Evidence from illusory crescents.” *Cognition* 170: 88–94. https://doi.org/10.1016/j.cognition.2017.08.007

Michel, Matthias. 2022a. “How (Not) to Underestimate Unconscious Perception.” *Mind & Language*, 1–18. https://doi.org/10.1111/mila.12406

Michel, Matthias. 2022b. “Confidence in Consciousness Research.” *WIREs Cognitive Science*, e1628. https://doi.org/10.1002/wcs.1628

Milner, David A. D., Monika Brechmann, and Luigi Pagliarini. 1992. “To Halve and to Halve Not: An Analysis of Line Bisection Judgements in Normal Subjects.” *Neuropsychologia* 30 (6): 515–526. https://doi.org/10.1016/0028-3932(92)90055-q

Morgan, Michael, Barbara Dillenburger, Sabine Raphael, and Joshua A. Solomon. 2012. “Observers Can Voluntarily Shift Their Psychometric Functions Without Losing Sensitivity.” *Attention, Perception & Psychophysics* 74 (1): 185–193. https://doi.org/10.3758/s13414-011-0222-7

Nelson, Thomas O. 1984. “A Comparison of Current Measures of the Accuracy of Feeling-of-Knowing Predictions.” *Psychological Bulletin* 95 (1): 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Peters, Megan A. K., Jeremy Fesi, Namema Amendi, Jeffrey D. Knotts, Hakwan Lau, and Tony Ro. 2017b. “Transcranial Magnetic Stimulation to Visual Cortex Induces Suboptimal Introspection.” *Cortex* 93: 119–132. 10.1016/j.cortex.2017.05.017

Peters, Megan A. K., Robert W. Kentridge, Ian Phillips, and Ned Block. 2017a. “Does Unconscious Perception Really Exist? Continuing The ASSC20 Debate.” *Neuroscience of Consciousness* 2017 (1): 1–11. https://doi.org/10.1093/nc/nix015

Peters, Megan A. K., and Hakwan Lau. 2015. “Human Observers Have Optimal Introspective Access to Perceptual Processes Even for Visually Masked Stimuli.” *ELife* 4 (October): 1–30. https://doi.org/10.7554/eLife.09651

Peters, Megan A. K., Tony Ro, and Hakwan Lau. 2016. "Who's Afraid of Response Bias?" *Neuroscience of Consciousness* 2016 (1): 1–16. https://doi.org/10.1093/nc/niw001

Phillips, Ian. 2016. "Consciousness and Criterion: On Block's Case for Unconscious Seeing." *Philosophy and Phenomenological Research* 93 (2): 419–451. https://doi.org/10.1111/phpr.12224

Phillips, Ian. 2018. "Unconscious Perception Reconsidered." *Analytic Philosophy* 59 (4): 471–514. https://doi.org/10.1111/phib.12135

Polat, Uri, and Dov Sagi. 2007. "The Relationship Between the Subjective and Objective Aspects of Visual Filling-In." *Vision Research* 47 (18): 2473–2481. https://doi.org/10.1016/j.visres.2007.06.007

Prinzmetal, William, Virginia Long, and James Leonhardt. 2008. "Involuntary attention and brightness contrast." *Perception & Psychophysics* 70 (7): 1139–1150. https://doi.org/10.3758/PP.70.7.1139

Rahnev, Dobromir. 2021. "A Robust Confidence–accuracy Dissociation via Criterion Attraction." *Neuroscience of Consciousness* 1 (2021): niab039. https://doi.org/10.1093/nc/niab039

Rahnev, Dobromir, Brian Maniscalco, Tashina Graves, Elliot Huang, Floris P. de Lange, and Hakwan Lau. 2011. "Attention Induces Conservative Subjective Biases in Visual Perception." *Nature Neuroscience* 14 (12): 1513–1515. https://doi.org/10.1038/nn.2948

Rajananda, Sivananda, Jeanette Zhu, and Megan A. K. Peters. 2018. "Normal Observers Show No Evidence for Blindsight in Facial Emotion Perception." *Neuroscience of Consciousness* 2020 (1): niaa023. https://doi.org/10.1093/nc/niaa023

Reingold, Eyal M., and Philip M. Merikle. 1988. "Using Direct and Indirect Measures to Study Perception without Awareness." *Perception and Psychophysics* 44 (6): 563–575. https://doi.org/10.3758/BF03207490

Reingold, Eyal M., and Philip M. Merikle. 1990. "On the Inter-Relatedness of Theory and Measurement in the Study of Unconscious Processes." *Mind & Language* 5 (1): 9–28. https://doi.org/10.1111/j.1468-0017.1990.tb00150.x

Sánchez-Fuenzalida, Gaal Nicolás, Van Simon, Stephen M. Fleming, and Julia M. Haaf. 2022. "Predictions and Rewards Affect Decision Making but Not Subjective Experience." https://doi.org/10.31234/osf.io/5v8dh

Samaha, Jason, Luca Iemi, and Bradley R. Postle. 2017. "Prestimulus Alpha-Band Power Biases Visual Discrimination Confidence, but Not Accuracy." *Consciousness & Cognition* 54 (2017): 47–55. https://doi.org/10.1016/j.concog.2017.02.005

Sekuler, Robert, and Albert Erlebacher. 1971. "The Invalidity of 'Invalid Results from the Method of Constant Stimuli': A Common Artifact in the Methods of Psychophysics." *Perception & Psychophysics* 9 (3): 309–311. https://doi.org/10.3758/BF03212655

Sekuler, Robert, Allison B. Sekuler, and Renee Lau. 1997. "Sound Alters Visual Motion Perception." *Nature* 385 (6614): 308. https://doi.org/10.1038/385308a0

Shams, Ladan. 2002. "Integration in the Brain-The Subconscious Alteration of Visual Perception by Cross-Modal Integration." *Science and Consciousness Review* 1: 1–4.

Shams, Ladan, Yukiyasu Kamitani, and Shinsuke Shimojo. 2000. "Illusions. What You See Is What You Hear." *Nature* 408 (6814): 788. https://doi.org/10.1038/35048669

Snodgrass, Michael, Edward Bernat, and Howard Shevrin. 2004. “Unconscious Perception: A Model-Based Approach to Method and Evidence.” *Perception and Psychophysics* 66 (5): 846–867. https://doi.org/10.3758/bf03194978

St. John-Saaltink, Elexa, Peter Kok, Hakwan C. Lau, and Floris P. de Lange. 2016. “Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex.” *Journal of Neuroscience* 36 (23): 6186–6192. https://doi.org/10.1523/JNEUROSCI.4390-15.2016

Swets, John A. 1973. “The Relative Operating Characteristic In Psychology.” *Science* 182 (4116): 990–1000. https://doi.org/10.1126/science.182.4116.990

Thurlow, Willard R., and Charles E. Jack. 1973. “Certain Determinants of the ‘Ventriloquism Effect’.” *Perceptual and Motor Skills* 36 (3): 1171–1184. https://doi.org/10.2466/pms.1973.36.3c.1171

Winawer, Jonathan, Alexander C. Huk, and Lera Boroditsky. 2008. “A Motion Aftereffect From Still Photographs Depicting Motion.” *Psychological Science* 19 (3): 276–283. https://doi.org/10.1111/j.1467-9280.2008.02080.x

Winawer, Jonathan, Alexander C. Huk, and Lera Boroditsky. 2010. “A Motion Aftereffect from Visual Imagery of Motion.” *Cognition* 114 (2): 276–284. https://doi.org/10.1016/j.cognition.2009.09.010

Witt, Jessica K., J. Eric T. Taylor, Mila Sugovic, and John T. Wixted. 2015. “Signal Detection Measures Cannot Distinguish Perceptual Biases from Response Biases.” *Perception* 44 (3): 289–300. https://doi.org/10.1068/p7908

Witthoft, Nathan, Long Sha, Jonathan Winawer, and Roozbeh Kiani. 2018. “Sensory and decision-making processes underlying perceptual adaptation.” *Journal of Vision* 18 (8): 1–20. https://doi.org/10.1167/18.8.10

Yarrow, Kielan, Nina Jahn, Szonya Durant, and Derek H. Arnold. 2011. “Shifts of Criteria or Neural Timing? The Assumptions Underlying Timing Perception Studies.” *Consciousness and Cognition* 20 (4): 1518–1531. https://doi.org/10.1016/j.concog.2011.07.003

Yeshurun, Yaffa, Maria Carrasco, and Laurence T. Maloney. 2008. “Bias and Sensitivity in Two-Interval Forced Choice Procedures: Tests of The Difference Model.” *Vision Research* 48 (17): 1837–1851. https://doi.org/10.1016/j.visres.2008.05.008