

Sensitivity to visual features in inattentional blindness

Reviewed Preprint

v2 • February 3, 2025

Revised by authors

Reviewed Preprint

v1 • October 4, 2024

Makaela Nartker , Chaz Firestone, Howard Egeth, Ian Phillips 

Department of Psychological & Brain Sciences, Johns Hopkins University; Baltimore, United States • Department of Philosophy, Johns Hopkins University; Baltimore, United States

 https://en.wikipedia.org/wiki/Open_access
 Copyright information

eLife Assessment

This study presents **valuable** findings to the field interested in inattentional blindness (IB), the phenomenon that participants fail to notice salient stimuli when their attention is directed elsewhere. This study reveals that participants who indicate no awareness of unexpected stimuli through yes/no questions ("did you notice anything unusual?"), may still show above-chance sensitivity to specific properties of these stimuli through follow-up forced-choice questions (e.g., regarding its location or color). By introducing absent trials where no IB stimulus is presented, the authors show that this is because participants are generally conservative and biased to report not noticing in inattentional blindness experiments. The evidence supporting these conclusions is **convincing**, the samples sizes are large and the analysis protocol is novel.

<https://doi.org/10.7554/eLife.100337.2.sa3>

Abstract

Summary

The relation between attention, perception and awareness is among the most fundamental problems in the science of the mind. One of the most striking and well-known phenomena bearing on this question is *inattentional blindness* (IB; Neisser & Becklen, 1975; Mack & Rock, 1998; Most et al., 2001, 2005). In IB, naïve observers fail to report clearly visible stimuli when their attention is otherwise engaged—famously even missing a gorilla parading before their eyes (Simons & Chabris, 1999). This phenomenon and the research programs it has motivated carry tremendous theoretical significance, both as crucial evidence that awareness requires attention (Cohen et al., 2012; Prinz, 2012; Noah & Mangun, 2020) and as a key tool in seeking the neural correlates of consciousness (Rees et al., 1999; Pitts et al., 2014; Hutchinson, 2019). However, these and other implications critically rest on a notoriously biased measure: asking participants whether they noticed anything unusual (and interpreting negative answers as reflecting a complete lack of perception). Here, in the largest ever set of IB studies, we show that, as a group, inattentionally blind participants can successfully report the location, color and shape of the stimuli they deny noticing. This residual visual sensitivity shows that perceptual information remains accessible in IB. We further show that subjective reports in IB are conservative, by introducing absent trials where no IB stimulus is presented; this

approach allows us to show for the first time that observers collectively show a systematic bias to report not noticing in IB—essentially ‘playing it safe’ in reporting their sensitivity. This pair of results is consistent with an alternative hypothesis about IB, namely that inattentionally blind subjects retain some degree of awareness of the stimuli they fail to report. Overall, these data provide the strongest evidence to date of significant residual visual sensitivity in IB. They also challenge the use of inattentional blindness to argue that awareness requires attention.

Main

One of the most pervasive and compelling intuitions about perception is that we see what is right in front of us: If a large enough stimulus were to appear directly before our eyes (with good lighting, a well-functioning sensory apparatus, no special camouflage, etc.), we would see it and could easily report features such as its color, shape, and location. However, this seemingly secure assumption is challenged by perhaps the best-known result in contemporary perception science: The phenomenon of inattentional blindness (IB). In IB, engaging in an attentionally demanding task (e.g., judging the relative lengths of two briefly presented lines, or counting basketball passes) causes observers to miss large, highly visible, but unexpected stimuli appearing right before their eyes (e.g., a high-contrast novel shape, or a parading gorilla; **Figure 1**). Whereas in ordinary circumstances these stimuli would be extremely salient and easily noticed, under conditions of inattention we seemingly do not see them at all.

IB is an extremely robust phenomenon, demonstrated in a stunning variety of laboratory and real-world contexts across half a century of research (Neisser & Becklen, 1975; Moore & Egeth, 1997; Mack & Rock, 1998; Simons & Chabris, 1999; Most et al., 2005; Drew et al., 2013; Murphy & Greene, 2016). Although the proportion of subjects demonstrating IB (the “IB rate”) varies widely depending on the protocol, it can be remarkably high (e.g., ~50% in the case of the famous gorilla, and over 80% in some of Mack & Rock’s studies; for reviews see Simons, 2000; De Brigard & Prinz, 2010; Jensen et al., 2011; Nobre et al., 2020; Redlich et al., 2021).

IB is also an extremely significant phenomenon. According to the “consensus view” (Noah & Mangun, 2020; see also Mack & Rock, 1998; Dehaene et al., 2006; Cohen et al., 2012; Prinz, 2012, 2015; though see Koch & Tsuchiya, 2007; Maier & Tsuchiya, 2020), subjects undergoing IB “have no awareness at all of the stimulus object” (Rock et al., 1992), such that “one can have one’s eyes focused on an object or event ... without seeing it at all” (Carruthers, 2015). This interpretation has endowed IB with numerous theoretical and practical implications. First, IB is a key piece of evidence for the more general claim that awareness requires attention (or, conversely, that “without attention, conscious perception cannot occur”; Dehaene et al., 2006), and so has been used to support various leading theories of consciousness on which attention plays a critical role, such as global neuronal workspace theory (Dehaene & Naccache, 2001), attended intermediate representation (AIR) theory (Prinz, 2015), and attention schema theory (Webb & Graziano, 2015). Second, conceived as a tool to abolish awareness, IB is frequently used by neuroscientists to measure neural activity in the absence of consciousness (Rees et al., 2000; Pitts et al., 2014; Hutchinson, 2019), in the hope of isolating the neural correlates of consciousness. Third, IB challenges cherished assumptions about our ability to perceive the world around us—inspiring substantial theoretical and philosophical debate (e.g., O’Regan & Noë, 2001; Lamme, 2003; Block, 2011; Cohen & Dennett, 2011) and rightly contributing to its status as one of the few results in perception science that has captured public interest and imagination (Chabris & Simons, 2011; Cloud, 2010; Murphy, 2017).

Crucially, however, this interpretation of IB and the many implications that follow from it rest on a measure that psychophysics has long recognized to be problematic: simply asking participants whether they noticed anything unusual. In IB studies, awareness of the unexpected stimulus (the

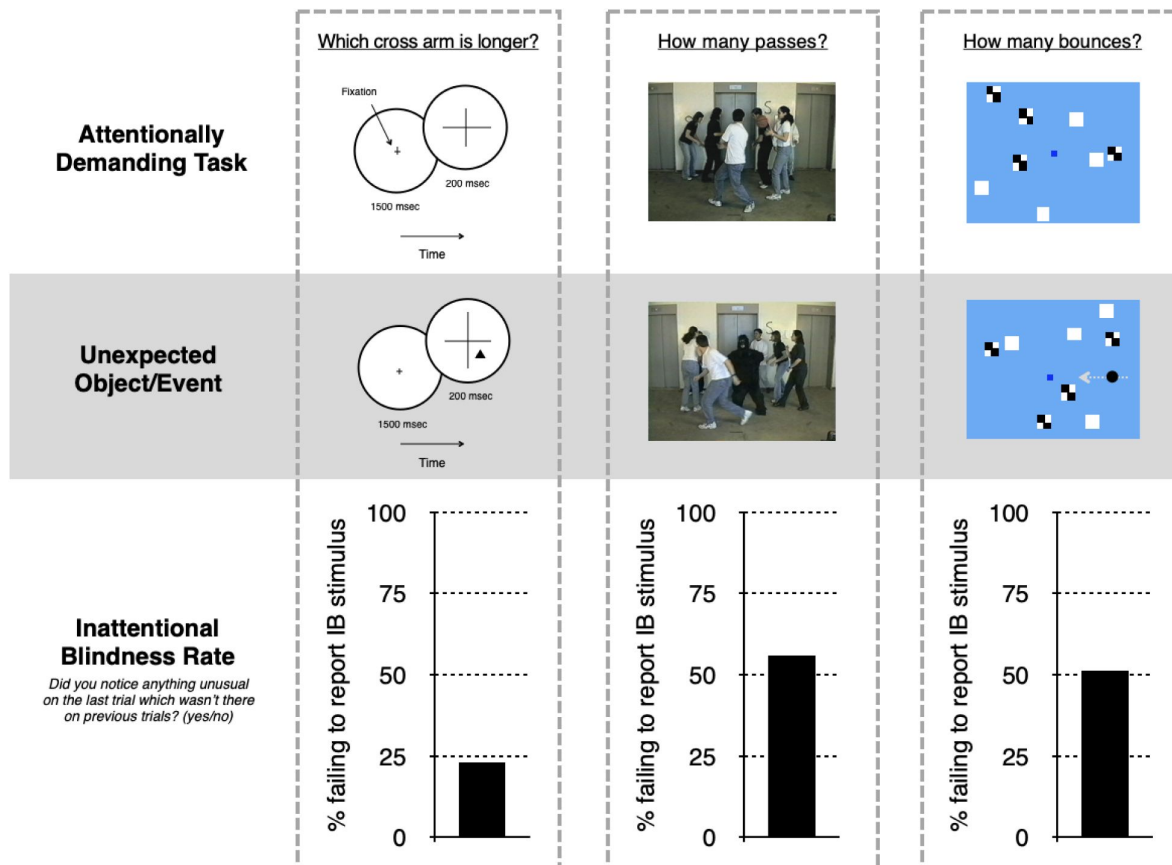


Figure 1.

Influential inattentional blindness paradigms and their results. (Left)

In a representative study from [Mack & Rock \(1998\)](#), subjects fixated in the center of the circle and reported which arm of a centrally presented cross was longer. On the critical trial, an unexpected shape appeared in the near periphery at the same time as the cross and subjects were asked whether they noticed anything unusual. **(Center)** In [Simons & Chabris \(1999\)](#), subjects counted the number of basketball passes made by individuals wearing white shirts. (Images drawn from [Simons and Chabris, 1999](#), reprinted with permission Dan Simons.) On the critical trial, a woman in a gorilla suit paraded through the display for five seconds, and subjects were asked whether they noticed anything unusual. **(Right)** In [Wood & Simons \(2017\)](#); based on [Most et al., 2001](#), subjects fixated on the blue square and reported the number of times the white or checkerboard squares bounced off the walls of the display. On the critical trial, a black circle entered from the right and crossed the display for several seconds, and subjects were asked whether they noticed anything unusual. In all three experiments (and many others in this literature), a considerable proportion of subjects reported not noticing these unexpected stimuli.

novel shape, the parading gorilla, etc.) is retroactively probed with a yes/no question, standardly, “Did you notice anything unusual on the last trial which wasn’t there on previous trials?”. Any subject who answers “no” is assumed not to have any awareness of the unexpected stimulus.

However, yes/no questions of this sort are inherently and notoriously subject to bias, because they require observers to set a *criterion* in order to decide whether they have enough evidence to answer “yes” or instead answer “no” (Dulany, 2001 [↗](#); cf. Eriksen, 1960 [↗](#); Holender, 1986 [↗](#); Irvine, 2012 [↗](#)). Under the framework of Signal Detection Theory (Tanner & Swets, 1954 [↗](#); Green & Swets, 1966 [↗](#)), observers who are asked to determine whether a signal is or is not present must in setting such a criterion consider tradeoffs between the various possible outcomes of their decision (i.e., the relative costs of hits, misses, false alarms, and correct rejections). Indeed, in an IB task (where the signal in question is the unexpected stimulus), observers may have reason to be conservative in their criterion-setting—i.e., to adopt a high standard for saying “yes” (and answer “no” if that standard isn’t met). For example, observers might be under-confident whether they saw anything (or whether what they saw counted as unusual); this might lead them to respond “no” out of an excess of caution. Subjects might doubt that they could identify it if asked, and so respond “no” to avoid having to do so. Subjects may also worry that if they report noticing the unexpected object, the experimenters will take that to mean they weren’t engaged in the task they had been given (e.g., judging which cross arm was longer or counting the passes; Dulany, 2001 [↗](#)). On any of these possibilities, subjects who say they did not notice the critical stimulus may well have had some awareness of it, but simply underreported it given the constraints of traditional IB questioning.

Evidence for this alternative hypothesis would have dramatic consequences for the dominant interpretation of IB and the implications that follow from it: It would overturn the view that IB reflects a total failure to perceive and so challenge the appeal to IB in arguments that attention is required for awareness. Remarkably, however, the hypothesis that subjects in inattentional blindness tasks actually see more than they say (but respond otherwise because they are conservative in reporting their awareness) remains empirically unsettled, and the powerful tools of signal detection theory unexploited in relation to IB.

A handful of prior studies have explored the possibility that inattentionally blind subjects may retain some visual sensitivity to features of IB stimuli (e.g., Schnuerch et al., 2016 [↗](#); see also Kreitz et al., 2020 [↗](#), Nobre et al., 2020 [↗](#)). However, a recent meta-analysis of this literature (Nobre et al., 2022 [↗](#)) argues that such work is problematic along a number of dimensions, including underpowered samples and evidence of publication bias that, when corrected for, eliminates effects revealed by earlier approaches, concluding “that more evidence, particularly from well-powered pre-registered experiments, is needed before solid conclusions can be drawn regarding implicit processing during inattentional blindness” (Nobre et al., 2022 [↗](#)).

Here, five experiments provide this critical test for the first time. We achieve this by making four key modifications to classic IB paradigms. First, we add follow-up questions to the classic yes/no methodology to explicitly probe awareness of features of the IB stimulus (i.e., not just asking whether subjects noticed anything, but also querying features of the object they said they didn’t notice). Although the use of follow-up questions is not new, our approach overcomes several problems with their past use in the literature (see General Discussion for much more on this point). Second, we include “absent trials” in which some subjects are shown no additional stimulus on the critical trial but are still asked whether they noticed anything unusual. This procedure provides a false alarm rate to test—for the first time, to our knowledge—whether subjects are conservative in reporting their awareness in the ways hypothesized above. Third, we leverage online data collection to massively increase sample size (running 25,000 subjects, each experiencing just a single critical trial), in order to make possible the crucial signal-detection analyses that separate bias from sensitivity. Finally, to overcome the fact that in IB each subject

necessarily only contributes one trial (since following a critical trial, stimuli are no longer unexpected), we introduce a novel analytic approach by applying signal detection models to a “super subject”.

Altogether this approach reveals that as a group subjects can report above-chance the features of stimuli (color, shape, and location) that they had all claimed not to notice under traditional yes/no questioning, and that this underreporting of awareness is well-modeled in terms of a conservative criterion. In place of the dominant interpretation that IB abolishes perception and awareness, the present results indicate that significant residual sensitivity remains in IB, and are even consistent with an alternative picture on which inattention instead *degrades* awareness. More generally, our findings motivate an approach to perception and awareness which treats them as graded as opposed to all-or-nothing, which we further discuss below.

Results

Above-chance sensitivity to inattentional blindness stimuli: Location

Experiment 1 modified the canonical inattentional blindness task used by Mack and Rock (1998; [Figure 2a](#)). On three trials, subjects were presented with a cross randomly assigned on each trial to be directly above or below a central fixation point (for 200 msec), and their task was to report which arm of the cross (horizontal or vertical) was longer. The fourth, critical trial, proceeded in the same way, but with the addition of an unexpected red line appearing in the periphery simultaneous with the cross. After again reporting which arm of the cross was longer, subjects were asked the standard question used to measure inattentional blindness: “Did you notice anything unusual on the last trial that wasn’t there on previous trials?” (yes or no). In line with established findings on IB, a considerable proportion of subjects (28.6%) responded “no” they didn’t notice anything unusual (we refer to these subjects across all our experiments as ‘non-noticers’).

Following the standard IB question, our modification included an additional question: “In fact, the last trial you just saw contained one extra element—a vertical red line on either the left or the right side of the box. What side do you think it appeared on?” (left or right). Far more subjects were able to correctly locate the stimulus (87.4%) than said they noticed it (71.4%), raising the possibility that non-noticers—i.e., those who demonstrated inattentional blindness—might have performed significantly better than chance. This is precisely what we found: 63.6% of non-noticers answered the forced-choice question correctly (significantly different from the 50% correct expected by chance in a two-sided binomial probability test, 95% CI = [0.54, 0.73], BF10 = 9.9). In other words, nearly two-thirds of subjects who had just claimed not to have noticed any additional stimulus were then able to correctly report its location.

Critically, this result also holds using d' , an unbiased measure of sensitivity ($d'_{2afc} = 0.51$, 95% CI = [0.16, 0.85]; see [Box 1](#) on Signal Detection Theory), which we use in reporting all following results concerning sensitivity in non-noticers. An important novelty of our strategy is that it derives these statistics in relation to a “super subject” whose responses are comprised of individual subjects’ responses in their single critical trials. Note that all analyses reported here relate to this super subject as opposed to individual subjects; see General Discussion for more on the assumptions behind this analytic approach.

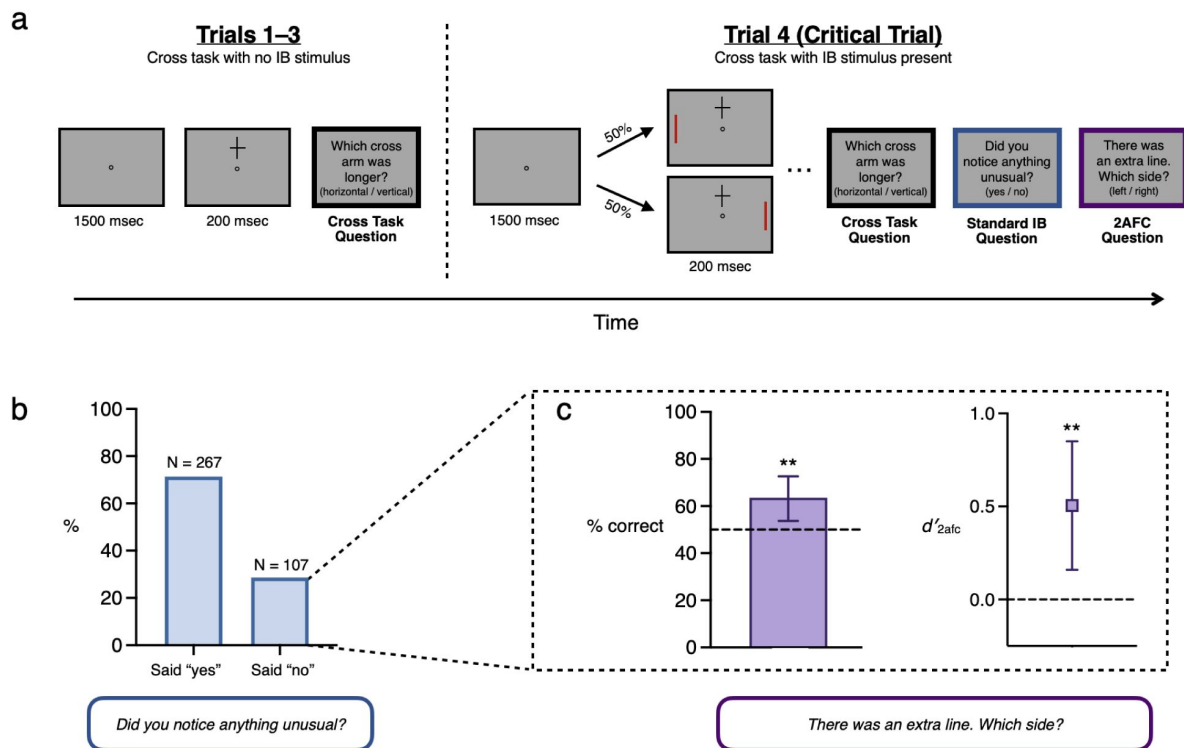


Figure 2.

Stimuli, procedure and results for Experiment 1.

a Schematic trial sequence for Experiment 1. On trials 1-3 subjects were presented with a cross above or below fixation for 200 msec and judged which arm was longer. On trial 4, an unexpected red line appeared in the periphery simultaneous with the cross. After reporting which cross-arm was longer, subjects were asked, "Did you notice anything unusual on the last trial that wasn't there on previous trials?" (yes/no), followed by a 2afc question concerning the location of the line (left/right). **b** Percentage of subjects who report noticing or not noticing the extra red line. 29% of subjects were 'non-noticers'. **c** Performance on the 2afc question (left/right location of the line), considering only those subjects who reported not noticing anything unusual. Remarkably, both %-correct and d' were significantly above chance among the group of subjects who met traditional criteria for inattention blindness. In other words, collectively, subjects who answered "no" demonstrated sensitivity to the location of the stimulus they all had just claimed not to have noticed. Error bars are 95% CIs.

Box 1: Signal Detection Theory and Inattentional Blindness

In Signal Detection Theory (SDT; [Green and Swets, 1966](#) [↗](#); [Macmillan and Creelman, 2005](#) [↗](#)), perceptual decisions are based on statistically variable sensory evidence ($S + N$) due to a stimulus or signal (S), if any, together with omnipresent noise (N). Deciding whether a stimulus was present or not (yes/no detection) involves deciding whether the evidence received is a sample from $S + N$ or from N alone. To make this decision, an observer must set a *criterion*, a level of sensory evidence sufficient for a positive response. This criterion is flexible and can be adjusted in accord with the payoffs associated with different outcomes such as hits (saying “yes” when a stimulus is present) and false alarms (saying “yes” when no stimulus is present).

SDT uses observed hit and false alarm rates to distinguish two distinct aspects of an observer’s performance: their *sensitivity* and *bias*. In simple models, N and $S + N$ distributions are treated as equal variance Gaussians. The *sensitivity* of an observer is then measured as the standardized distance between the means of the N and $S + N$ distributions. Intuitively, if the two distributions entirely overlap, an observer is entirely insensitive to the presence of a stimulus, whereas the greater the separation, the greater the sensitivity. To calculate this measure of sensitivity, we subtract the z-transform of the false alarm rate from the z-transform of the hit rate: $d' = z(H) - z(FA)$. Critically, this measure is independent of the location of a subject’s criterion and so offers an objective, bias-free measure of their sensitivity.

The *bias* of an observer is their tendency to prefer a particular response independent of the actual presence of a stimulus. This can be measured by the location of their criterion with respect to the midpoint of the N and $S + N$ distributions. Intuitively, a criterion at the midpoint shows no preference for “yes” responses over “no” responses independent of the actual presence of a stimulus, whereas a *conservative* criterion reflects a preference for “no” responses, and a *liberal* criterion a preference for “yes” responses. We can calculate a standardized measure of bias, as follows: $c = -\frac{1}{2}[z(H) + z(FA)]$.

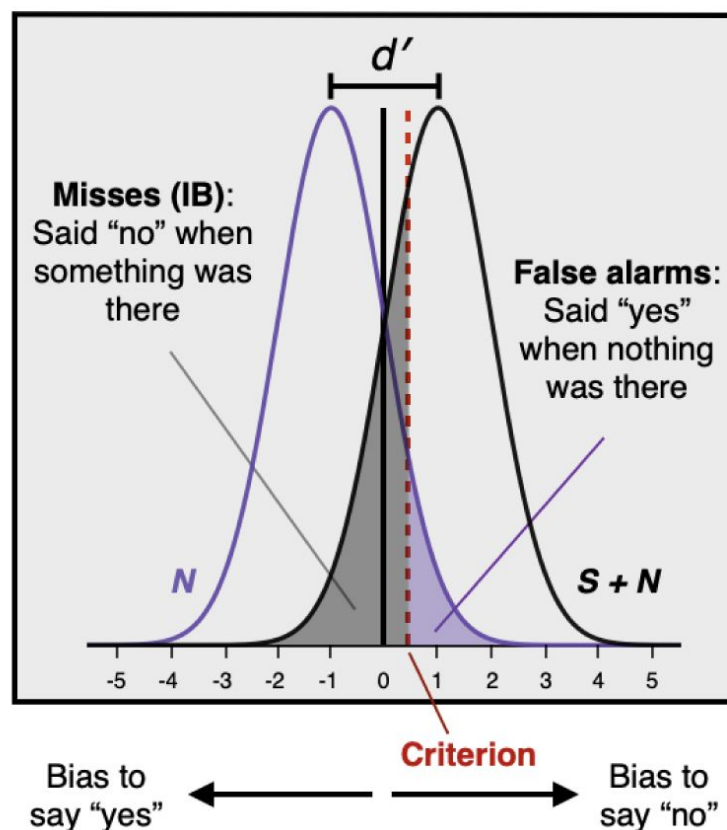
The traditional question in IB studies—“Did you notice anything unusual on the last trial that wasn’t there on previous trials?”—can be treated as a yes/no detection question. By including absent trials, where no stimulus was presented and yet we ask this question anyway, we determined hit and false alarm rates across critical trials after applying the [Hautus \(1995\)](#) [↗](#) correction of adding 0.5 to every cell in the decision matrix to avoid infinite values which would arise if any cell were zero. For example, in our Exp. 5, 7.2% of subjects who were shown no IB stimulus nonetheless responded “yes”—yielding a corrected false alarm rate of 0.073. On the other hand, 71% of subjects who were shown a stimulus responded “yes”—yielding a corrected hit rate of 0.71. Using the calculations described, we could separate the sensitivity of the corresponding super-subject from their bias. This revealed a significant *conservative* bias ($c = 0.45$), indicating a preference to deny noticing an unexpected stimulus, independent of its actual presence.

“Did you notice anything unusual?”

	“Yes”	“No”
Present	5730	2339
Absent	200	2561

$$\text{Corrected hit rate (say “Yes” when Present)} = \frac{5730 + 0.5}{(5730 + 0.5) + (2339 + 0.5)} = 0.71$$

$$\text{Corrected false alarm rate (say “Yes” when Absent)} = \frac{200 + 0.5}{(200 + 0.5) + (2561 + 0.5)} = 0.073$$



Above-chance sensitivity to inattentional blindness stimuli: Color

Experiment 2 repeated the design of Experiment 1, except that the additional line could be either red or blue, and the question about the line’s location was replaced with a one-interval forced-response question about the line’s color. This experiment yielded similar results to Experiment 1: Only 72.3% of subjects shown an additional stimulus said they had noticed something, yet 81% were able to indicate the additional line’s color correctly. And again, as a group, non-noticers demonstrated above chance sensitivity, this time to color ($d' = 0.38$, 95% CI = [0.03, 0.73])—note that the unbiased nature of this measure is especially critical here since subjects displayed a significant bias in favor of responding “blue” ($c = 0.67$, 95% CI = [0.49, 0.84]). In other words, consistent with the results of Experiment 1, as a group, subjects who had just claimed not to have noticed an

additional stimulus were able to correctly report its color at rates well above chance. This pair of initial results shows that even subjects who answer “no” under traditional questioning can, as a group, still correctly report various features of the stimuli they just reported not having noticed, indicating significant group-level sensitivity to visual features. Moreover, these results are even consistent with an alternative hypothesis about IB, that as a group, subjects who would traditionally be classified as inattentionally blind are in fact at least partially aware of the stimuli they deny noticing.

Conservative reporting of visual awareness

Our results raise a natural question: Why, if subjects could succeed at our forced-response questions as a group, did they all individually claim not to have noticed anything? Experiment 2 made an additional modification precisely to address this question: the introduction of ‘absent’ trials in which no additional line was shown but subjects were still asked the yes/no and one-interval forced-response questions. Absent trials provide an additional source of information about subjects’ biases, by revealing how often they respond “yes” without any stimulus present (i.e., their false alarm rate). This allows for the computation of response bias (c) in relation to the crucial IB question (“Did you notice anything unusual on the last trial that wasn’t there on previous trials?” again, see [Box 1](#)), which to our knowledge is unique in this literature. This analysis revealed evidence for the second aspect of our alternative hypothesis: Not only do subjects collectively have residual sensitivity to unnoticed IB stimuli, they are, as a group, *conservative* in reporting their awareness ($c = 0.31$, 95% CI = [0.22, 0.41]; $d' = 1.81$, 95% CI = [1.63, 1.99]; note that statistics here are for all subjects). In turn, this raises the possibility that subjects may retain awareness of unreported IB stimuli corresponding to their residual visual sensitivity but that this awareness is systematically underreported.

Above-chance sensitivity in high-confidence non-noticers

Although the previous two studies are consistent with the hypothesis that “inattentionally blind” subjects retain at least partial awareness of unattended stimuli (and are conservative in reporting that awareness), it is possible that these results were driven by a subset of subjects, with other subjects remaining truly blind to the IB stimulus (i.e., having no sensitivity at all to its features). This might arise if above-chance sensitivity were restricted to subjects who were under-confident in responding “no” when asked whether they had noticed anything unusual. In that case, subjects who *confidently* answered “no” (i.e., felt certain they didn’t notice any additional stimulus) might, as a group, fail to perform above chance on the subsequent discrimination task. Experiment 3 addressed this possibility directly, by (a) adding confidence ratings to the standard yes/no question, and (b) dramatically increasing the sample size, so as to separately analyze the performance of high- and low-confidence subjects. The task proceeded in the same way as Experiment 1, except that after the yes/no question about noticing anything unusual, subjects were asked to rate their confidence in their answer, on a 4-point scale from 0–3 (0 = Not at all confident; 3 = Highly confident); finally, subjects were then asked the left/right discrimination question (and gave their confidence in that answer as well, though this was less crucial to our hypothesis—see Supplementary Materials for details). As shown in [Figure 3e](#), answers to these questions ran the full spectrum of responses, with subjects expressing varying degrees of confidence in “yes” and “no” responses to the IB question. Of particular relevance was the group of “high-confidence non-noticers”—i.e., subjects who said “no” (they didn’t notice anything unusual), and then rated their confidence in that answer as “3” (highly confident). Remarkably, even this group of subjects ($N = 204$) collectively demonstrated significantly above-chance sensitivity to the location of the IB stimulus ($d'_{2afc} = 0.34$; 95% CI = [0.08, 0.60]). Further, as is evident from [Figure 3e](#), subjects’ confidence in their yes/no response predicted accuracy for that group on the discrimination task, suggesting that subjects may have graded awareness of unattended stimuli in IB tasks (see Supplementary Materials for details).

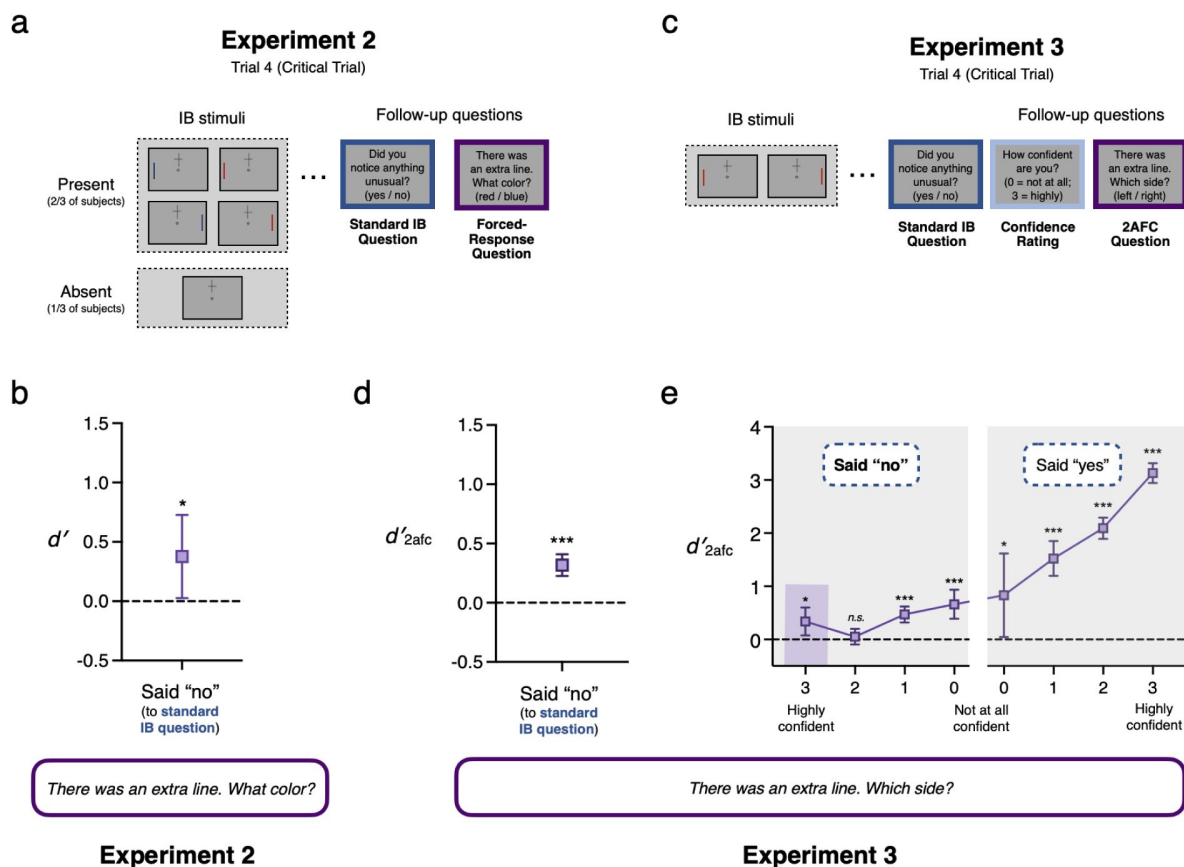


Figure 3.

Stimuli, procedure and results for Experiments 2 and 3.

a Schematic trial sequence for Experiment 2. On trials 1-3 subjects were presented with a cross above or below fixation for 200 msec and judged which arm was longer. On trial 4, for 2/3rds of subjects, an unexpected line appeared in the periphery simultaneous with the cross. This line could be either blue or red and one the left or right. For 1/3 of subjects no additional line was shown. After reporting which cross-arm was longer, all subjects were asked, "Did you notice anything unusual on the last trial that wasn't there on previous trials?" (yes/no), followed by a one-interval forced-response question concerning the color of the line (red/blue). **b** Performance on the one-interval forced-response question about the unexpected line's color amongst subjects who were shown a line and who reported not noticing anything unusual (27.73% of subjects). As a group, subjects who answered "no" demonstrated sensitivity to the color of the stimulus they had just claimed not to have noticed. **c** Schematic trial sequence for Experiment 3. Trials 1-4 were identical to Experiment 1, except subjects were asked additional questions about their confidence following both yes/no and 2afc questions. **d** Performance on the 2afc question in Experiment 3, considering only subjects who reported not noticing anything unusual (30.85% of subjects). Replicating the finding of Experiment 1, as a group, subjects who answered "no" demonstrated sensitivity to the location of the stimulus they had just claimed not to have noticed. **e** Performance on the 2afc question in Experiment 3 for all subjects, broken down by confidence in their response to the yes/no question whether they had noticed anything unusual. Remarkably, even subjects who were highly confident that they had not noticed anything unusual were collectively significantly above chance. Error bars are 95% CIs.

Generalizing to dynamic inattentional blindness

Experiments 1-3 suggest that, collectively, subjects underreport their perception of a brief (200 msec) IB stimulus. However, in classic studies of dynamic inattentional blindness (Simons & Chabris, 1999 [↗](#); Most et al., 2005 [↗](#); Ward & Scholl, 2015 [↗](#)), the unexpected stimulus remains in view for an extended period. Here, it is tempting to think that subjects will not be conservative in reporting that they noticed an IB stimulus given they have many seconds to build confidence in what they saw. Experiment 4 tested this possibility by modifying a traditional sustained inattentional blindness paradigm in which the IB stimulus remains on screen for 5 seconds.

Finding the same pattern of above-chance sensitivity and conservative response bias in this very different paradigm would be striking evidence that, quite generally, there is residual sensitivity to visual features in inattentional blindness, and lend further support to our alternative hypothesis.

In Experiment 4, subjects' primary task was to count how often squares of a particular color (black or white) bounced off the perimeter of a gray rectangle (adapted from Wood & Simons, 2017 [↗](#), in turn based on Most et al., 2001 [↗](#); see **Figure 4a** [↗](#)). This task demands significant attention as each set of colored squares bounces an average of 28 times during a 17-second trial. For some subjects, on the third and critical trial, an additional shape (a triangle or circle, which was either black or white) entered the display (on the left or the right) and traversed the full height of the display in a straight path (from either top to bottom, or bottom to top), remaining on screen for 5 seconds before disappearing. When the trial ended, subjects reported how many times the squares of their assigned color bounced. They were then immediately asked the standard question used to measure inattentional blindness: "Did you notice anything unusual on the last trial that wasn't there on previous trials?" (yes or no). Following this, each subject answered *three* additional questions about the extra object that may or may not have appeared, in a random order: (1) What color was it? (black or white), (2) What shape was it? (circle or triangle), (3) What side was it on? (left or right).

Consistent with previous studies using dynamic stimuli (see **Figure 1b** [↗](#) and **1c** [↗](#)), a majority of subjects—57.3%—demonstrated IB. Put another way, just 42.7% of subjects who were shown something additional on the critical trial were correct on the yes/no detection question. However, despite the attentionally demanding primary task, non-noticers collectively again demonstrated significant sensitivity to the features of the IB stimulus, choosing the correct color ($d' = 0.82$, 95% CI = [0.61, 1.04]) and shape ($d' = 0.21$, 95% CI = [0.01, 0.42]) significantly more often than would be predicted by chance. Given that all the objects were changing location constantly in Experiment 4, we did not pre-register a prediction about location discrimination being above-chance, and we did not find that non-noticers performed above chance on the left/right discrimination question (though we did observe a trend in this direction, $d'_{2afc} = 0.07$, 95% CI = [-0.08, 0.22]).

Just as in Experiment 2, subjects underreported their awareness of the IB stimulus on the traditional yes/no question, employing a decision criterion that was even more conservative than that measured in Experiment 2 ($c = 0.85$, 95% CI = [0.73, 0.97]; $d' = 1.33$, 95% CI = [1.10, 1.57]; again, note that statistics here are for all subjects). The fact that subjects were collectively more reluctant to report their awareness of the IB stimulus in this experiment compared with Experiment 2 is surprising given that subjects had 5 full seconds to build confidence that there was a stimulus. However, given that the colors of the squares in the primary task were the same as the colors used for the IB stimuli, it is possible that estimates of criterion and color sensitivity in Experiment 4 were affected by an interaction effect due to enhancement of color-congruent stimuli and suppression of color-incongruent stimuli (cf. Wood & Simons, 2017 [↗](#)). Specifically, the IB rate for color-incongruent stimuli was substantially higher than for color-congruent stimuli (86.57% vs. 28.71%), suggesting suppression occurred. That is, a subject shown an unexpected black stimulus while attending to black squares was more likely to report noticing than a subject shown an

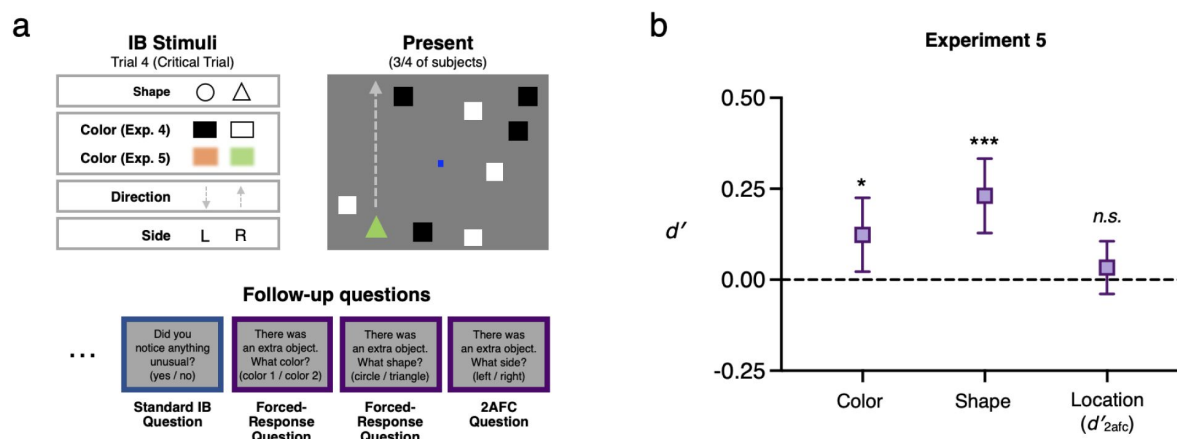


Figure 4.

Stimuli, procedure and results for Experiment 5.

a Stimulus parameters and a schematic critical trial for Experiment 5, in which an unexpected object moved across the display for several seconds while subjects counted the number of times the white squares bounced off the display's 'walls' (task adapted from Wood & Simons, 2017). The unexpected object varied in its shape, color, direction of motion, and side of the display. As in Experiment 4, subjects were then asked follow-questions about this stimulus's color, shape and location.

b As a group, subjects who reported not noticing the unexpected stimulus still showed above-chance sensitivity to its color and shape (though not to its location), a pattern predicted in our pre-registration. Thus, sensitivity to IB stimuli arises even when the stimuli are visible for a sustained period (rather than appearing only briefly, as in Experiment 1-3). Error bars are 95% CIs.

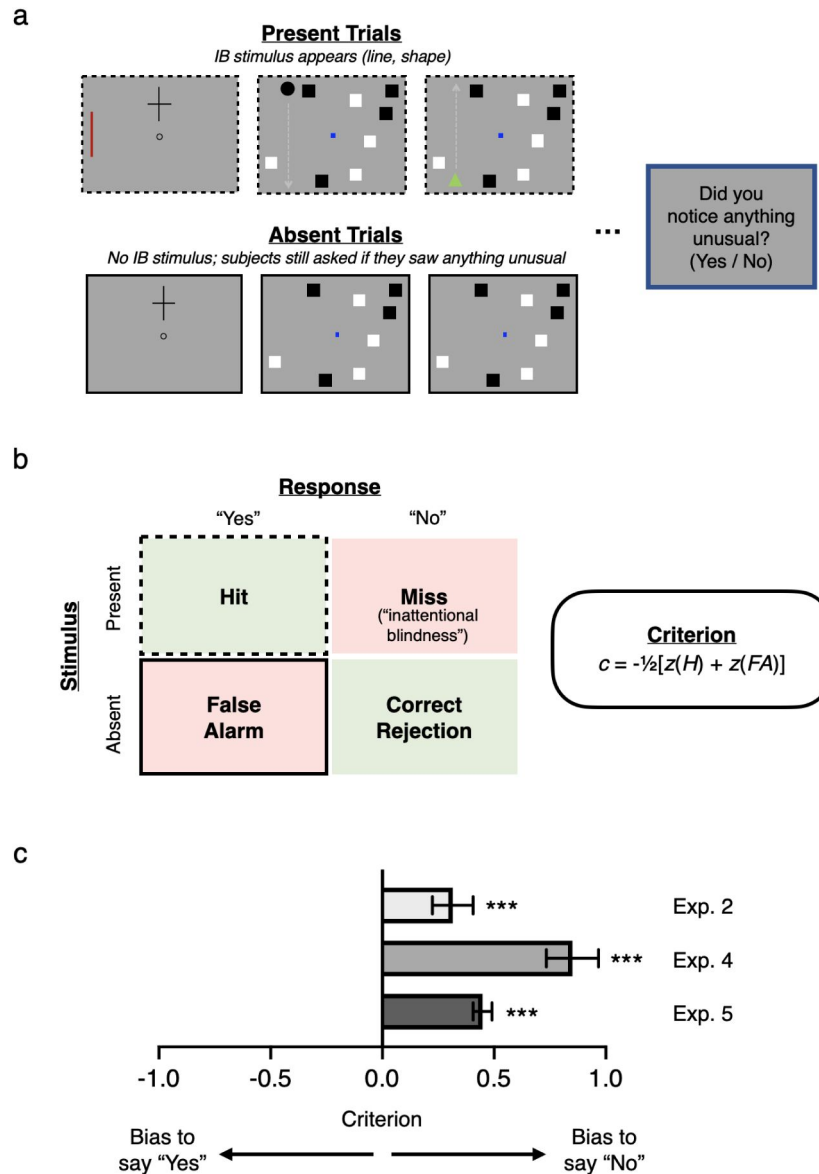


Figure 5.

Conservative criteria in Experiments 2, 4 and 5.

a Critical trials from Experiments 2, 4 and 5, showing present trials in which an IB stimulus was presented to subjects (2/3rds subjects in Experiment 2, 3/4 in Experiments 4 and 5), and absent trials in which no IB stimulus was presented (1/3rd subjects in Experiment 2, 1/4 in Experiments 4 and 5). On all trials, subjects were asked, "Did you notice anything unusual on the last trial that wasn't there on previous trials?" (yes/no). **b** Left: Decision matrix for this yes/no question, indicating the four possible stimulus/response pairings: hits corresponding to 'yes' responses on present trials; false alarms corresponding to 'yes' responses on absent trials; misses corresponding to 'no' responses on present trials (i.e. inattentional blindness); and correct rejections corresponding to 'no' responses on absent trials. Right: Formula for calculating criterion or response bias from hit rate (H), i.e. proportion of present trials in which subjects responded 'yes', and false alarm rate (FA), i.e. proportion of absent trials in which subjects responded 'no'. **c** Criteria for Experiments 2, 4 and 5, showing that in each case subjects exhibited a significantly conservative bias, i.e. a tendency to say 'no' when asked if they noticed anything unusual, independent of the actual presence of a stimulus. This suggests that, collectively, subjects in inattentional blindness experiments may systematically underreport their awareness of unexpected stimuli across different paradigms.

unexpected white stimulus. This in turn may have inflated our estimate of color sensitivity if subjects correctly assumed that they would have been more likely to notice a color-congruent stimulus and biased their responses towards the incongruent color.

To address this, Experiment 5 repeated the design of Experiment 4, with two key differences: (1) Every subject was assigned to attend to the white squares in the primary task, (2) the IB stimuli were either orange or green—two colors that as a pair produced equal IB rates in pilot testing. Moreover, by selecting such highly salient colors (which differ dramatically from any other stimulus on the display), these parameters served as an especially strong test of our broader hypothesis. Remarkably, even under these conditions (and with the concern about color congruency effects mitigated), the pattern of results matched that of Experiment 4: Subjects collectively performed above-chance in discriminations of both color ($d' = 0.12$, 95% CI = [0.02, 0.23]) and shape ($d' = 0.23$, 95% CI = [0.13, 0.33]), but not location ($d'_{2afc} = 0.03$, 95% CI = [-0.04, 0.11]). Sensitivity to the color of the IB stimulus was lower than in Experiment 4, consistent with our prediction that color sensitivity in that experiment was inflated by a color-congruence bias; however, it was still significantly above chance, indicating that color-congruence does not fully account for the pattern of results. That subjects, as a group, could still correctly report the color and shape of the IB stimulus despite the attentionally demanding primary task is compelling evidence that perception of IB stimuli is not completely abolished by inattention. Further, we again found that subjects were collectively biased to respond “no” to the traditional yes/no IB question ($c = 0.45$, 95% CI = [0.41, 0.49]; $d' = 2.01$, 95% CI = [1.94, 2.09]; statistics here are for all subjects). Thus, even when an unexpected stimulus remains on screen for many seconds, subjects are hesitant to report noticing anything unusual.

General Discussion

Inattention blindness captures both scholarly interest and popular imagination because of its striking and counterintuitive implication that we may fail to see what is right before our eyes, simply because our attention is otherwise engaged. Its influence is both wide and deep: It apparently provides a dramatic demonstration of the limits of visual perception, serves as a tool to reveal the neural correlates of consciousness, and even motivates theories of consciousness holding that awareness requires attention. These and still other implications arise from the “consensus” interpretation of IB, according to which inattention completely abolishes perception of the unexpected stimulus: “one can have one’s eyes focused on an object or event ... without seeing it *at all*” (Carruthers, 2015 [\[link\]](#), emphasis added). Yet this interpretation rests on a crucial and untested assumption: that observers who say they didn’t notice a stimulus (i.e., answer “no” under traditional yes/no questioning) in fact didn’t see it.

The present work puts this crucial assumption to the test, yielding results that point to a very different pattern than the consensus interpretation: Across 5 pre-registered experiments totaling over 25,000 subjects, we found that groups of observers could successfully report a variety of features of unattended stimuli, even when they all individually claimed not to have noticed those stimuli. Furthermore, our approach revealed that subjects are conservatively biased in reporting their awareness, in ways that not only explain our results (i.e., provide an account of how and why subjects who claimed not to have seen something could still report its features) but also recast the large and influential body of literature that has taken answers to yes/no questions in IB paradigms at face value.

Design and analytical approach

Our experiments all employed designs and protocols closely modeled on canonical IB studies. In Experiments 1-3, we studied IB using a cross task closely modeled on Mack and Rock’s classic studies (Mack & Rock, 1998 [\[link\]](#)). In Experiments 4-5, we studied IB using a dynamic task closely

modeled on Wood & Simons (2017) (itself adapted from influential work by Most et al., 2001; Most et al., 2005; Ward & Scholl, 2015). In all cases, we used the standard yes/no question from previous experiments to determine IB rates. These choices were deliberate: Our aim was to interrogate the canonical interpretation of a large and long-standing tradition of experimental work, and so we sought to cleave as closely as possible to the experiments which inspired and have been subject to this interpretation. Our results do not reflect idiosyncratic design choices but rather speak to the central paradigms in the literature.

Our approach also eschews the use of divided and full attention trials. Divided attention trials (trials after the critical trial in which the subject knows that an unexpected stimulus might occur) and full attention trials (where in addition the subject is told to ignore the primary task, and instead look out for unexpected objects) are often used to exclude from analysis subjects who do not report seeing the unexpected stimulus on these trials (see, e.g. Most et al. 2001). However, this practice is controversial and has been argued to “lead researchers to understate the pervasiveness of inattentional blindness” (White et al., 2018). Because our aim was to offer an especially stringent test of sensitivity in inattentional blindness, we opted not to use any such exclusions. Since the use of such exclusions would tend to inflate the biases we are concerned with here, the fact that we found evidence of residual sensitivity despite not excluding subjects in this way is all the more telling.

To assess objective sensitivity and bias in IB, we adopted a novel analytical approach, applying signal detection models to a “super subject” whose responses are comprised of individual subjects’ responses in their single critical trials. This strategy involves various assumptions which future work should explore. Nonetheless, we believe any violations of such assumptions will not affect our main results or, where they might, lead only to our underestimating residual sensitivity and response bias. First, in calculating d' , it is standardly assumed that a subject’s criterion is stable throughout a given experiment. This assumption may be more likely to be violated with respect to a “super subject”. However, its violation will not affect calculations of d' in 2afc tasks which are criterion free (Exps. 1, 3 and location results in Exps. 4 & 5). Moreover, if criterion instability is present, its effect will be to reduce estimated sensitivity in one-interval forced-response tasks (Exp. 2, and shape and color results in Exps. 4 & 5; see, Azzopardi & Cowey, 2001). Our approach can thus be seen as offering a conservative estimate of residual sensitivity. (Indeed, given the essentially retrospective nature of IB judgments, our estimates should in any case be considered conservative since signal available at the time of stimulus presentation may have been lost by the time of judgment.) Second, in calculating d' and c , it is standardly assumed that signal and noise distributions are equal variance Gaussians. There is theoretical reason to think this assumption is robust with respect to 2afc tasks (Macmillan & Creelman, 2005), and in general in relation to the Gaussian nature of the distributions (Green & Swets, 1966; Pastore et al., 2003; Wixted, 2020). However, empirically, equal variance might not hold in one-interval tasks. Violation of this equal variance assumption could lead to under- or over-estimation of d' and c . However, any such under- or over-estimation would be slight and unlikely to affect our main results. For example, if the relative variance of the signal-plus-noise distribution as compared to the noise distribution (σ) is 1.25, then c for the yes/no task in Experiment 5 (illustrated in Box 1) would be 0.443 and d' would be 1.898; whereas, if $\sigma = 0.75$, then c would instead be 0.438 and d' would be 2.118. In either case, our main results are qualitatively unchanged.

Our super subject analysis raises the question of how to interpret the responses of individual subjects. Even though our experiments revealed that subjects who denied noticing any unusual stimulus could *collectively* report its features above chance, it was also the case that some subjects denied noticing any unusual stimulus and then also went on to incorrectly answer the follow-up questions about its features. Indeed, this has also been true in other studies that include follow-ups (e.g., Most et al., 2001, 2005; Cohen et al., 2011, 2020). Are these individual subjects truly inattentionally blind? Intriguingly, even this seemingly cautious conclusion does not follow from that pattern of performance, for several reasons. First, and straightforwardly, many such follow-

ups are not 2afc questions but rather yes/no questions themselves (including our own questions about color and shape, as well as many similar questions used in previous work); since such questions are biased (e.g., subjects may tend to favor responding that an object was blue rather than red, perhaps making assumptions about the visibility of each color), any individual incorrect answer to such a question may reflect this sort of bias as opposed to the total absence of color signal. Second, even in an unbiased 2afc task, an observer may have significant information from the stimulus (perhaps, well above an unbiased single-interval detection criterion) but still decide incorrectly because of high noise from the other spatial interval; in such a case, it is far from clear that the subject should be treated as blind. Third, there will inevitably be subjects who fail to correctly report an unnoticed object's features because they failed to see it in the first place — not due to inattention, but rather due to more ordinary failures such as happening to look away from the display at the key moment, sneezing or blinking just as the unexpected stimulus appears, being interrupted by one's child or pet or smoke alarm, and so on; such subjects would be “blind” to the stimulus, of course, but not *inattentionally* blind. Fourth any series of follow-up questions, including ours, inevitably probes only some limited set of features at the exclusion of others; thus, subjects may have been aware of some feature of the stimulus other than the features explicitly probed (e.g., the orientation of an unexpected line rather than its color). Fifth, many processes intervene between being (or not being) sensitive to a stimulus and generating a response to a follow-up question; subjects will occasionally press the wrong button by mistake, or rush through the questions without reading carefully, or forget what they saw (and thus guess); though such mishaps will tend to break in favor of incorrect answers just as often as correct answers, they make it so that any individual error (or success) is difficult to interpret on its own (and in ways that testify to the value of the group-level approach that we favor here). Sixth, and perhaps most generally, taking correct and incorrect answers to place subjects neatly into two categories — those who saw the stimulus and those who did not — reflects a binary approach to perception and awareness that we suggest should be resisted. Indeed, an aspect of our contribution here, discussed further below, is to encourage conceiving perception and awareness as coming in degrees, in line with a Signal Detection Theory framework. On this view, perception and awareness are most helpfully characterized in terms of continuous statistics such as d' rather than more traditional but problematic measures such as the proportion of correct or incorrect responses.

Relation to previous work

Our work was motivated by concerns that the traditional interpretation of IB relies on assessing perception and awareness simply by asking participants whether they noticed anything unusual. As discussed, such yes/no questions are notoriously subject to bias, which may lead subjects to answer “no” even when they do have a degree of perception or awareness, due to factors such as underconfidence. More recent studies of IB have attempted to improve on simple yes/no questioning through the use of various follow-up questions. However, although these improved methods undoubtedly have their merits, none of them resolves the concerns that motivated our investigations. This is for three fundamental reasons.

First, follow up questions are often used not to *exclude* subjects from the IB group but to *include* subjects. For example, [Most et al. 2001](#) [↗](#) treated as inattentionally blind not only subjects who denied noticing the unexpected object but also subjects who claimed they did notice the object but were subsequently unable to describe it. Similarly, [Pitts et al. \(2012\)](#) [↗](#) asked subjects to rate their confidence in their initial yes/no response from 1 = least confident to 5 = most confident, and used these ratings to include in the IB group those who rated their confidence in seeing at 3 or less. Counting such subjects as inattentionally blind can be problematic. There is a large gap between being under confident that one saw something and being completely blind to it; failure to describe a feature (e.g., color, shape) does not imply a complete lack of information concerning that feature; and even if a subject did lack all information concerning some feature of an object, this would not imply a complete failure to see the object.

Second, most follow up questions remain subject to response bias in just the same way as the original yes/no awareness question. For example, [Cohen et al. \(2020\)](#); see similarly: [Cohen et al., 2011](#); [Simons et al., 1999](#); [Most et al., 2001](#), [2005](#); [Drew et al., 2013](#); [Memmert, 2014](#)) use a series of follow up probes: (1) “Did you notice anything strange or different about that last trial?” (2) “If I were to tell you that we did something odd on the last trial, would you have a guess as to what we did?” (3) “If I were to tell you we did something different in the second half of the last trial, would you have a guess as to what we did?” (4) “Did you notice anything different about the colors in the last scene?” These questions are intrinsically subject to response bias just by being yes/no questions; but in this case they may be especially susceptible since subjects may be reluctant to ‘take back’ their earlier answers, leaving them all the more conservative to avoid any perceived inconsistency. (This may also explain the remarkable consistency in such responses reported in, e.g., [Simons & Chabris, 1999](#), despite the very different wording across the questions asked.) It is also important to recognize that whereas 2afc questions are criterion free (in that they naturally have an unbiased decision rule), this is not generally true of *n*-alternative match to sample designs. Performance in such tasks thus requires SDT analysis – which itself may be problematic if the decision space is not properly understood or requires making substantial assumptions about observer strategy.

Third, and finally, many follow up questions are insufficiently sensitive (especially with small sample sizes). For instance, [Todd, Fougner & Marois \(2005\)](#) used a 12-alternative match-to-sample task (see similarly: [Fougner & Marois, 2007](#); [Devue et al., 2009](#)). And [Most et al. \(2005\)](#) asked an open-response follow-up: “If you did see something on the last trial that had not been present during the first two trials, what color was it? If you did not see something, please guess.” These questions are more difficult and to that extent less sensitive than binary forced-response/2afc questions of the sort we use in our own studies – a difference which may be critical in uncovering degraded perceptual sensitivity. For all these reasons, we believe our novel approach of using 2afc or forced-response questions combined with signal detection analysis is an important improvement on prior methods.

Previous studies of the related phenomenon of change blindness have investigated whether subjects who fail to detect changes nonetheless perform above chance in discrimination tasks concerning the changed object ([Mitroff et al., 2002](#), [2004](#); [Hollingworth & Henderson, 2002](#)). However, only a handful of prior studies have explored the possibility that inattentionally blind subjects outperform chance in reporting or responding to features of IB stimuli (e.g., [Schuerch et al., 2016](#); see also [Kreitz et al., 2020](#), [Nobre et al., 2020](#)). Moreover, a recent meta-analysis of this literature ([Nobre et al., 2022](#)) concluded that such work is problematic along a number of dimensions, including underpowered samples and evidence of publication bias that, when corrected for, eliminates effects revealed by earlier approaches. (These concerns hold in addition to our own worries about biased measures of performance.) The authors of this meta-analysis conclude with the following recommendation for future work: “We suggest that more evidence, particularly from well-powered pre-registered experiments, is needed before solid conclusions can be drawn regarding implicit processing during inattentional blindness” ([Nobre et al., 2022](#)). We see the present set of high-powered pre-registered studies as providing precisely this evidence, in ways that advance our understanding of IB considerably.

Our results also shed new light on evidence that inattentionally blind subjects process the unexpected stimuli they deny noticing. For example, in the electrophysiology literature, unexpected line patterns have been found to elicit the same Nd1 ERP component in both noticers and inattentionally blind subjects ([Pitts et al., 2012](#)). Likewise, preserved neural signatures of scene segmentation and perceptual inference (e.g., Kanizsa figures) have been found in inattentionally blind subjects (see respectively, [Scholte et al., 2006](#), and [Vandenbroucke et al., 2014](#)), though interestingly not differential responses to meaningful words versus meaningless letter strings ([Rees et al., 1999](#)). Similarly, behavioral studies show that unattended stimuli can influence the accuracy and speed of inattentionally blind subjects’ judgments in a primary task

(e.g., [Moore & Egeth, 1997](#); replicated in [Wood & Simons, 2019](#); [Pugnaghi et al., 2020](#)). Although some researchers have interpreted these results as implying a kind of subliminal processing of IB stimuli, our results just as easily raise an alternative (and perhaps more straightforward) explanation: that inattentionally blind subjects may retain a degree of awareness of these stimuli after all.

We acknowledge that above-chance performance in our experiments could be taken to reflect unconscious representations in a manner akin to orthodox interpretations of blindsight ([Weiskrantz et al., 1974](#); [Kolb & Braun, 1995](#); though see [Phillips, 2021a](#); [Morgan et al., 1997](#); and for more general skepticism, [Newell & Shanks, 2023](#)). However, in our view, explicit voluntary judgments of stimulus features (especially in neurotypical subjects) constitute *prima facie* evidence of conscious processing, and should be interpreted that way unless there is some compelling reason to favor an alternative ([Snodgrass, 2002](#); [Baldon & Azzopardi, 2015](#); [Heeks & Azzopardi, 2015](#); [Phillips, 2021b](#)). As a result, although we acknowledge the possibility that our data reflect unconscious processing, we think there are compelling reasons to interpret our results in terms of residual conscious vision in IB (although we note that this claim is tentative and secondary to our primary finding).

Evidence that inattentionally blind subjects process and are sensitive to the unexpected stimuli they deny noticing has been used to support so-called *inattentional amnesia* accounts of IB—the traditional rival to the orthodox interpretation of IB. On inattentional amnesia accounts, unattended objects and features are consciously perceived but not encoded so as to be available for later explicit report ([Wolfe, 1999](#); [Moore, 2001](#); though see [Ward & Scholl, 2015](#); [Hirschhorn et al., 2024](#)). Our results are consistent with some degree of inattentional amnesia, and likewise, with what [Block \(2001\)](#) calls *inattentional inaccessibility*. (For a fuller discussion of these alternative hypotheses and IB more generally, see [Wu, 2014](#).) However, our findings suggest that inattentional amnesia cannot be the whole story, since they reveal that some features of unexpected objects are available for later explicit report even in the group of subjects who deny noticing anything unusual.

Visual awareness as graded

A further upshot of our findings is that they lend support to a more graded perspective on IB (in particular) and both perception and visual awareness (in general). This stands in contrast to the two interpretations that have dominated discussion of IB, both of which adopt a binary perspective. On the orthodox interpretation, inattention abolishes all perception and awareness; on the rival inattentional amnesia account, inattention abolishes all explicit encoding. Our data suggest the need to move beyond such binaries (cf. [Cohen et al., 2023](#)): Inattention degrades but does not eliminate perception and awareness – and likewise explicit encoding. This more nuanced approach has some kinship with what [Simons \(2000\)](#) calls *inattentional agnosia*. On this account, subjects who report not noticing may have some awareness of the unexpected object but fail to “encode the properties necessary to register that the item was something new, different, or noteworthy” ([Most et al., 2005](#)). Though aligned with the spirit of our view, this account still does not fully capture our results, since in our studies the group of non-noticing subjects could explicitly report features which would ordinarily suffice to mark the unexpected object as new or different (e.g., color, shape). Nonetheless, we agree that unattended stimuli are encoded in a partial or degraded way. Here we see a variety of promising options for future work to investigate. One is that unattended stimuli are only encoded as part of ensemble representations or summary scene statistics ([Rosenholtz, 2011](#); [Cohen et al., 2016](#)). Another is that only certain basic “low-level” or “preattentive” features (see [Wolfe & Utochkin, 2019](#) for discussion) can enter awareness without attention. A final possibility consistent with the present data is that observers can in principle perceive individual objects and higher-level features under inattention but that the precision of the corresponding representations is severely reduced. Our central aim here is to

provide evidence that there is residual perceptual sensitivity to visual features for subjects who would ordinarily be classified as inattentionally blind. Further work is needed to characterize the exact nature of this perception, and the awareness (if any) which corresponds to it.

Conclusion

Taken together, and after decades of inconclusive findings, the results of our five studies offer the strongest evidence so far of significant residual visual sensitivity across a range of visual features in IB. In other words, as a group, the inattentionally blind enjoy at least some degraded or partial sensitivity to the location, color and shape of stimuli which they report not noticing. Together with our finding that subjects collectively exhibit a systematically conservative bias in reporting their awareness, our results also call into question the orthodox interpretation of IB on which inattention entirely abolishes awareness, suggesting that a reconceptualization of inattentional blindness may be required. Indeed, perhaps ironically, inattentional blindness if anything provides evidence that awareness of certain features *survives* inattention. Our results highlight the critical value of assessing response bias and including objective measures of sensitivity in studying inattentional blindness and visual awareness. They also point to a broader rethinking of perception and consciousness as graded, rather than binary, phenomena.

Methods

Open Science Practices

All sample sizes, exclusion criteria, analyses and key experimental parameters reported here have been pre-registered. Data, analyses, stimuli and pre-registrations are publicly available at <https://osf.io/fcrhu/>. Readers can also experience all experiments for themselves at <https://perceptionresearch.org/ib/>.

Experiment 1: Above-chance sensitivity to the location of unnoticed stimuli

Participants

500 adults were recruited from the online platform Prolific (for validation of the reliability of this subject pool, see [Peer et al., 2017](#)), with participation limited to US subjects. As described in our pre-registration, we reached this number by running batches of 100 subjects until a target number of 100 non-noticers (i.e., subjects answering “no” to the yes/no question about whether they noticed the unexpected stimulus) was reached. After excluding subjects who incorrectly reported which arm of the cross was longer on any of trials 1-3 and those who failed to provide a complete dataset, or failed a test for color vision (Ishihara color plate; see data archive), 374 subjects were included in the analysis. (All of these exclusion criteria were pre-registered.) This experiment and all others reported here were approved by the Homewood Institutional Review Board of Johns Hopkins University. All subjects provided informed consent and were compensated financially for their participation.

Stimuli and Procedure

As shown in [Figure 2](#), Experiment 1 contained four trials, with the fourth trial differing from the first three in several ways. All trials took place in a display with dimensions 600 px x 600 px. Due to the nature of online experiments, we cannot be sure of the exact size or distance of stimuli as subjects actually experienced them (and so we give these figures in pixels); however, any differences in subjects’ monitors and/or display setups would have been constant across all trials of the experiment.

On trials 1-3, a fixation circle appeared in the center of the display; subjects pressed the spacebar when they were ready to begin the trial. The keypress was followed by a 1500 msec delay, after which a cross formed by two thin black lines (3 px thickness; 200 px x 140 px dimensions) appeared for 200 msec either 150 px above or below the fixation circle. The location of the cross (above or below), as well as its aspect ratio (vertical-longer or horizontal-longer) was randomly chosen on each of these trials. The cross then disappeared, followed by a 500 msec blank interval. Subjects were then asked which of the cross's arms was longer (horizontal or vertical).

The fourth, critical trial proceeded the same way, except that simultaneous with the cross, a vertical red line (200 px long; 3 px thick; RGB(147,0,0)) appeared on one side of the display (10 px from the boundary of the display), also for 200 msec. Subjects were then asked the same cross arm length question as before.

Following this, two additional questions were asked, in the following order, each on its own display (such that subjects only saw the second question after answering the first).

Question 1

Did you notice anything unusual on the last trial which wasn't there on previous trials? (yes / no)

Question 2

In fact, the last trial you just saw contained one extra element — a vertical red line on either the left or the right side of the box. What side do you think it appeared on? If you don't know, or don't think it appeared on either side, take your best guess. (left / right)

For all forced-choice and forced-response questions asked in all experiments reported here, subjects indicated their answers by clicking a radio button next to the text corresponding to their answer, and then submitted their responses by clicking a separate “Submit” button, a design aimed at eliminating motor-error responses.

Analysis and Results

As reported in the main text, 28.6% (107/374) of subjects responded “no” to Question 1; we refer to these subjects as “non-noticers”. These subjects are those who demonstrate inattentional blindness by conventional standards. However, 63.6% of non-noticers answered Question 2 correctly (the 2afc location task). This proportion was compared to chance responding (50%) with both frequentist and Bayesian null hypothesis tests, using the `binom.test` and `proportionBF` functions, respectively, from the BayesFactor package in R (Ihaka et al., 1996 [↗](#); Morey et al., 2015 [↗](#); all arguments used were defaults), which yielded a 95% CI = [0.54, 0.73], and a BF10 = 9.9.

SDT analyses began by calculating the number of hits, false alarms, present trials, and absent trials, and then applying the log-linear correction of adding 0.5 to all cells of the decision matrix (Hautus, 1995 [↗](#); see also Box 1 [↗](#)). This is a standard correction to prevent an infinite d' in the event that either hits or false alarms are zero (this correction was applied to all experiments, though neither hits nor false alarms were ever zero), and simulations have shown that if anything, this correction *underestimates* d' (Hautus, 1995 [↗](#)).

The signal detection theory measure of sensitivity (d') for non-noticers' performance on the 2afc location task was calculated as follows:

$$d'_{2afc} = \frac{1}{\sqrt{2}} [z(H) - z(FA)]$$

For our analysis, we (arbitrarily) considered trials where a stimulus was presented on the *left* to be “present” trials, and trials where a stimulus was presented on the *right* to be “absent” trials (d' will be identical regardless of which trial type is considered present/absent; c will be the same value with the opposite sign). With a hit rate and false alarm rate of 72.64% and 45.54% respectively (after log-linear adjustment), the resulting $d'_{2afc} = 0.51$. Note that d' for this experiment was adjusted downward by a factor of 1/2 because 2afc tasks are theoretically easier than yes/no or forced-response tasks (Macmillan and Creelman, 2005 [\[1\]](#)). This formula for d'_{2afc} (as well as those for d' and *criterion*, described below) assumes equal variance for signal and noise distributions (a standard assumption in SDT), but the analysis code provided allows for the calculation of SDT statistics when the variance of signal and noise distributions is unequal.

Because each subject in the experiment contributes just one trial, the signal detection metrics were calculated at the group level, and the standard error for each SDT calculation was estimated using methods described by Macmillan and Creelman (2005 [\[1\]](#), pgs. 325-328; see also Kadlec, 1999 [\[2\]](#)). We estimated the variance for d' in this 2afc task using methods first described by Gourevitch and Galanter (1967) [\[3\]](#), and re-described in Macmillan and Creelman (2005) [\[1\]](#), equations 13.5 and 13.7. First, equation 13.5 demonstrates how ϕ (a function which converts z-scores into probabilities) can be computed for the hit rate and false alarm rates:

$$\phi(H) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z(H)^2}$$

$$\phi(FA) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z(FA)^2}$$

With $\phi(H)$ and $\phi(FA)$ computed, we then estimated the variance of d' in this 2afc task using equation 13.7:

$$\text{var}(d'_{2afc}) = \frac{H(1-H)}{2N_2[\phi(H)]^2} + \frac{FA(1-FA)}{2N_1[\phi(FA)]^2}$$

where N_2 is the number of present trials, and N_1 is the number of absent trials. (For information about how the variance is affected by sample size using different methods and in different tasks, see Macmillan and Creelman Table 13.2 and 13.3; our sample sizes are more than sufficient to expect this variance estimation to be accurate to the hundredth decimal place).

Finally, we computed a confidence interval around d'_{2afc} using standard methods: The result is 95% CI = [0.16, 0.85], suggesting performance in the non-noticing group was significantly above chance.

Experiment 2: Above-chance sensitivity to the color of unnoticed stimuli

Participants

1700 adults were recruited from Prolific, collected in batches of 100 subjects until a target number of 100 non-noticers was reached. After excluding subjects who incorrectly reported which arm of the cross was longer on any of trials 1-3 and those who failed to provide a complete dataset, or failed a test for color vision, 1261 subjects were included in the analysis.

Stimuli and Procedure

The fourth, critical trial proceeded the same way as Experiment 1, except that the extra vertical line that appeared simultaneous with the cross was either red (RGB(147,0,0)) or blue (RGB(0,0,136)), with the color and location of the line randomized across subjects.

Following the presentation of Trial 4, subjects were asked which cross arm was longer, and then the same traditional IB question as in Experiment 1 (Question 1), followed by:

Question 2

The last trial you just saw contained one extra element — a vertical line on one side of the box. What color was the extra line? If you don't know, or don't think any line appeared, take your best guess. (red / blue)

To reduce uncertainty about what color “red” and “blue” referred to, the text for each color option was printed in the red and blue color used for the IB stimuli in the different conditions (RGB(147,0,0), RGB(0,0,136)).

Analyses and Results

As reported in the main text, 27.7% of subjects shown an additional stimulus responded “no” to Question 1 (i.e., demonstrated inattentive blindness by conventional standards). However, 58.5% of these non-noticers answered correctly on Question 2 (95% CI = [51.95%, 64.93%]; BF10 = 4.54).

In Experiment 2, the follow-up color discrimination was a two-alternative forced-response design, and so for signal detection analyses, sensitivity was calculated without the $1/\sqrt{2}$ adjustment, such that

$$d' = z(H) - z(FA)$$

resulting in a $d' = 0.38$. We estimated the variance of d' for this one-interval forced-response task using similar methods to those described for Experiment 1, with one minor change to the variance equation to account for this being a forced-response task (equation 13.4 in [Macmillan and Creelman, 2005](#)):

$$\text{var}(d') = \frac{H(1-H)}{N_z[\phi(H)]^2} + \frac{FA(1-FA)}{N_1[\phi(FA)]^2}$$

The only difference between equation 13.4 and equation 13.7 is that the latter includes a factor of 2 in the denominator of both terms, which accounts for 2afc tasks theoretically being easier than yes/no tasks.

The procedure for significance testing comparing d' to chance ($d' = 0$) was identical to Experiment 1 Methods, and the two-sided frequentist binomial probability test yielded a 95% confidence interval of [0.03, 0.73]. As stated in the main text, subjects demonstrated a significant bias to respond “blue,” which we measured by calculating subjects’ criterion on the red/blue question as,

$$c = -\frac{1}{2}[z(H) + z(FA)]$$

resulting in a positive (conservative) criterion value of $c = 0.67$. The 95% confidence interval around the criterion estimate is [0.49, 0.84]. Since this interval does not contain zero, this represents a statistically significant bias. For a more intuitive understanding of the bias, 78.42% of subjects shown *no IB stimulus* guessed that the stimulus was blue.

Lastly, an important contribution of this work is the inclusion of absent trials, which enable us to compute response bias (c) for the traditional IB question, “Did you notice anything unusual on the last trial that wasn’t there on previous trials?” As predicted, we found that, as a group, subjects were conservative in reporting their awareness of the IB stimulus ($c = 0.31$, 95% CI = [0.22, 0.41]), suggesting that subjects in inattentive blindness experiments may systematically underreport their awareness of unexpected stimuli.

Experiment 3: Above-chance sensitivity even in highly confident non-noticers

Participants

7,000 subjects were recruited from Prolific, and data were collected in batches until we reached 200 non-noticers who reported being “highly confident” in their yes/no response. After excluding duplicate data files, subjects who incorrectly reported which arm of the cross was longer on any of trials 1-3, those who failed to provide a complete dataset, or failed a test for color vision, 5296 subjects were included in the analysis.

Stimuli and Procedure

This experiment explored whether sensitivity to the visual features of an IB stimulus varies as a function of subjects’ confidence in their responses to the traditional IB question (i.e., whether or not they noticed anything unusual on the last trial), and more specifically whether even highly confident non-noticers would show such sensitivity. Experiment 3 was identical to Experiment 1, except that after subjects answered the traditional yes/no question (which was the same as Question 1 in Experiment 1) and the follow-up left/right question (which was the same as Question 2 in Experiment 2), they also rated their confidence in each of those responses:

Question 1

Did you notice anything unusual on the last trial which wasn’t there on previous trials? (yes / no)

Question 2 (Confidence rating for Question 1)

How confident are you in your answer? (Four-point scale: 0 = Not at all confident – 3 = Highly confident)

Question 3

The last trial you just saw contained one extra element - a vertical red line on either the left or the right side of the box. What side do you think it appeared on? If you don’t know, or don’t think it appeared on either side, just take your best guess. (left / right)

Question 4 (confidence rating for Question 3; not shown in [Figure 3](#))

How confident are you in your answer? (Four-point scale: 0 = Not at all confident – 3 = Highly confident)

Analyses and Results

As reported in the main text, 30.85% of subjects shown an additional stimulus responded “no” to Question 1 (i.e., demonstrated inattentional blindness by conventional standards)

As with Experiments 1 and 2, we were interested in whether subjects who responded “no” to the traditional IB question would nevertheless perform above-chance on subsequent discrimination questions about the IB stimulus. Beyond our general interest in all subjects who answered “no,” we were most interested in whether the non-noticers who reported high confidence in their answer still performed above-chance on the 2afc question. We compared the performance of those observers to chance ($d' = 0$) using the methods described above, and found that even the most confident non-noticers—i.e., those who reported being “highly confident” that they did not notice

anything unusual, rating their confidence at 3 on a scale from 0 to 3—collectively demonstrated significantly above-chance sensitivity to the location of the IB stimulus: $d'_{2afc} = 0.34$; 95% CI = [0.08, 0.60]. Importantly, this group of subjects is minimally powered, with just 204 subjects, adding force to the argument that there is meaningful sensitivity amongst highly confident non-noticers. The bin of second-most interest is that of the moderately confident non-noticers (those responding with a confidence rating of 2 on a scale from 0 to 3). Sensitivity in this group was above zero ($d'_{2afc} = 0.05$), but not significantly so (95% CI = [-0.10, 0.20]).

Figure 3e [↗](#) depicts the uncorrected results of null hypothesis significance tests comparing the d'_{2afc} estimate for each confidence rating bin to chance ($d'_{2afc} = 0$), but essentially the same pattern of significant results is obtained when the Holm-Bonferroni correction for multiple comparisons is applied ($p = .034$ for the No-3 bin, and $p = 0.09$ for the Yes-0 bin—though note this bin only contains 25 subjects).

Experiment 4: Above-chance sensitivity in a sustained inattentional blindness task

Participants

1500 subjects were recruited from Prolific, and data were collected in batches until we reached 100 non-noticers who answered the color discrimination question first, 100 non-noticers who answered the shape discrimination question first, and 100 non-noticers who answered the location discrimination question first (more on these three discrimination questions below).

After exclusions, 1278 subjects including 417 non-noticers remained in the analysis for Experiment 4. The pre-registered exclusion criteria for this experiment were the same as those reported by [Wood & Simons \(2017\)](#) [↗](#). Subjects were excluded if: (i) their reported bounces for either of the first two trials erred by more than 50% in either direction from the true number of bounces of their attended objects on that trial, (ii) they failed to contribute a complete dataset, (iii) they reported problems with experimental playback, such as stuttering, freezing or another issue specified in a free-response, or (iv) an observer managed to submit or run the study twice (evidenced by two files sharing the same Prolific ID), in which case we excluded their second run from the analysis.

Stimuli and Procedure

In Experiment 4, subjects participated in a sustained inattentional blindness task modified slightly from experimental code published by [Wood & Simons \(2017\)](#) [↗](#). All subjects completed three trials of a dynamic, multiple object tracking task containing black squares and white squares. At the beginning of the experiment, each subject was told whether they should attend to the black or white squares. At the beginning of each trial, subjects were instructed to fixate on a small blue square (11 px x 11 px; RGB(0,0,255)) in the center of a gray (RGB(127,127,127)) rectangle (666 px x 546 px), and were told that their task was to count the number of times the squares of their attended color bounced off of the walls of the gray rectangle. Each trial lasted approximately 17 seconds, and each subset of black/white squares bounced an average of 28 times. After the trial ended, subjects were asked to report how many times the squares of their attended color bounced off of the walls of the rectangle. Of the 222 excluded subjects, 185 were excluded for bounce reports that erred by more than 50% in either direction of the actual number of bounces.

The critical trial was Trial 3. In this trial, 3/4 of subjects (Present condition) were shown an unexpected shape (a circle or a triangle, which was either black or white). This unexpected object entered the display on either the left or the right 5 seconds after the trial began, and moved either

upward or downward until it exited the other side of the display (thus, there were 2 color x 2 shape x 2 side x 2 motion direction options for IB stimuli, randomly chosen for each subject). For 1/4 of subjects, no additional object was shown on Trial 3 (Absent condition).

Regardless of condition, at the end of the critical trial, subjects were again asked to report the number of bounces, followed by four additional questions:

Question 1

Did you notice something on the last trial that did not appear on previous trials? (yes/no)

After answering Question 1 (the standard IB question), subjects were told (again regardless of condition): *An extra object may have appeared on that last trial. If you saw it, please tell us its color, shape, and whether it appeared on the left or the right side of this gray box. If you didn't see the extra object, please guess. We'll ask about [whichever discrimination question was randomly chosen to be asked first] first.*

Questions 1-3 (presented in random order for each subject)

The new object was... (black/white)

The new object was a... (circle/triangle) The new object was on the... (left/right)

Note that Question 1 (the traditional IB question) differs slightly in Experiments 4 and 5 from the question wording in Experiments 1-3 because we aimed to cleave as closely as possible to [Wood & Simons \(2017\)](#) and other inattention blindness experiments using similar paradigms.

Analyses and Results

As reported in the main text, 57.32% of subjects shown an additional stimulus on the critical trial answered “no” to the traditional IB question (Question 1). For absent trials, 6.31% of subjects answered “yes” to the traditional IB question when no additional stimulus appeared; this is the false alarm rate, which can be used (along with the hit rate) to estimate subjects' bias in responding to the traditional IB question. In Experiment 4, as in Experiment 2, we found that, as a group, subjects answered the traditional IB question using a conservative criterion ($c = 0.85$, 95% CI = [0.73, 0.97]), suggesting they may be underreporting their awareness of IB stimuli.

Experiment 4 asked subjects to perform 3 discrimination tasks following their response to Question 1, with the order randomized for each subject. The color and shape discriminations are one-interval forced response tasks, and so d' was calculated for these two questions without the 2afc correction. We found that non-noticers collectively performed significantly above chance on both the color ($d' = 0.82$, 95% CI = [0.61, 1.04]) and shape ($d' = 0.21$, 95% CI = [0.005, 0.42]) discriminations. In addition to the issue identified below, the fact that this result was marginal further motivated Experiment 5, in which we substantially increased the number of subjects recruited in order to reduce the size of the confidence intervals around d' estimates of discrimination sensitivity by 50%. For the location discrimination (left/right), the 2afc correction to d' was applied, and non-noticers' performance was trending but was not above-chance ($d'_{2afc} = 0.07$, 95% CI = [-0.08, 0.22]).

Because the order of the discrimination questions varied by subject, we pre-registered an analysis specifying that we would also derive sensitivity and bias for subjects who answered a given discrimination question first. With these subsets of non-noticers (roughly one-third the original sample size), confidence intervals are much larger, and although color discrimination remained

significantly above-chance ($d' = 0.78$, 95% CI = [0.41, 1.16]; $N = 189$), shape discrimination was no longer significant ($d' = 0.15$, 95% CI = [-0.20, 0.50]; $N = 195$) and location ($d'_{2afc} = -0.17$, 95% CI = [-0.44, 0.10]; $N = 176$) was again not significant.

Finally, in a pre-registered analysis breaking down the inattentional blindness rate by the color of the attended stimuli, we found that subjects were roughly 3x more likely to report noticing a color-congruent than a color-incongruent IB stimulus (86.57% vs. 28.71%). This interacted with subjects' responses on the color discrimination task, with subjects shown an IB stimulus demonstrating a significant bias to say that the stimulus that appeared was the opposite color to the colored squares they attended (84.25% of subjects who attended to white squares answered black; 75.09% of subjects who attended to black squares answered white). In order to get a better estimate of sensitivity to color in this task, we pilot tested a new pair of IB stimulus colors in Experiment 5 to equate the guessing rate on absent trials.

Experiment 5: Replicating above-chance sensitivity in a sustained inattentional blindness task

Participants

To ensure a large enough sample without overlap with previous experiments using this paradigm, Experiment 5 recruited Prolific subjects not only from the USA but also from Canada, the United Kingdom, and Australia. In order to reduce the size of the confidence intervals around the sensitivity estimates relative to Experiment 4, we collected data until we reached at least 2,200 subjects who reported not noticing the unexpected stimulus. Exclusion criteria were identical to Experiment 4. After exclusions, 10,830 subjects in total, including 2,339 non-noticers, were included in the analysis. (To our knowledge, this made Experiment 5 the largest single inattentional blindness sample ever collected.)

Stimuli and Procedure

Experiment 5 repeated the design of Experiment 4, with two key differences: (1) Every subject was assigned to attend to the white squares in the primary task, (2) the IB stimuli were either orange or green—two colors that as a pair produced equal IB rates in pilot testing. Again, subjects were asked the traditional IB question (yes/no) followed by three discrimination questions (color, shape, and location) in random order for each subject. The change to the colors of the IB stimuli of course meant that the color question and options were changed to read:

The new object was... (color 1/color 2; with the text printed in the given color)

Analyses and Results

As reported in the main text, with the congruency effects mitigated and despite the highly salient color of the orange/green IB stimuli, the pattern of results matched that of Experiment 4: Analysis of responses to the traditional yes/no question once again revealed that subjects were collectively biased to respond “no” ($c = 0.45$, 95% CI = [0.41, 0.49]), with the hit rate (subjects in the Present condition who responded “yes”) being 71.01% (100% - the IB rate) and the false alarm rate (subjects in the Absent condition who responded “yes”) being 7.24% (see a more detailed breakdown of the SDT analysis for this experiment in [Box 1](#)). Thus, even when a highly salient, moving stimulus entered the display suddenly and remained on screen for multiple seconds, subjects were hesitant to report noticing anything unusual.

Non-noticers' performance on the three discrimination questions was also consistent with the pattern of results in Experiment 4, with non-noticers collectively performing significantly above-chance in discriminating both color ($d' = 0.12$, 95% CI = [0.02, 0.23]) and shape ($d' = 0.23$, 95% CI =

[0.13, 0.33]), but not location ($d'_{2afc} = 0.03$, 95% CI = [-0.04, 0.11]). Biases in the discrimination question responses among subjects shown no IB stimulus on the critical trial (the Absent condition) were minimal, confirming that the congruency effect (interaction between the IB rate and subjects' choices on the color discrimination question) was mitigated in Experiment 5: For subjects shown no unexpected stimulus on the critical trial (the Absent condition), 52.95% guessed that the extra object had been orange, and 47.05% guessed that it had been green. On the shape question, 51.43% of subjects guessed that the extra object had been a triangle, and 48.57% guessed that it had been a circle. Finally, on the location question, 54.87% of subjects guessed that the extra object had been on the left side of the display, and 45.13% guessed that it had been on the right side of the display.

References

- Azzopardi P., Cowey A, De Gelder B., De Haan E. H. F., Heywood C. A. 2001) **Why is blindsight blind?** *Out of mind: Varieties of unconscious processes* Oxford University Press :3–19
- Balsdon T., Azzopardi P 2015) **Absolute and relative blindsight** *Consciousness and Cognition* **32**:79–91
- Barbot A., Xue S., Carrasco M 2021) **Asymmetries in visual acuity around the visual field** *Journal of Vision* **21**:1–23
- Block N 2001) **Paradox and Cross Purposes in Recent Work on Consciousness** *Cognition* **79**:197–219
- Block N 2011) **Perceptual consciousness overflows cognitive access** *Trends in Cognitive Sciences* **15**:567–575
- Carruthers P 2015) **The centered mind: What the science of working memory shows us about the nature of human thought** Oxford: OUP
- Chabris C., Simons D 2011) **The invisible gorilla: How our intuitions deceive us.** Harmony
- Cloud J. 2010) **Human Perception and the Invisible Gorilla** *TIME* <https://content.time.com/time/health/article/0,8599,2003097,00.html>
- Cohen M. A., Alvarez G. A., Nakayama K 2011) **Natural-scene perception requires attention** *Psychological Science* **22**:1165–1172
- Cohen M. A., Botch T. L., Robertson C. E 2020) **The limits of color awareness during active, real-world vision** *Proceedings of the National Academy of Sciences* **117**:13821–13827
- Cohen M. A., Cavanagh P., Chun M. M., Nakayama K 2012) **The attentional requirements of consciousness** *Trends in Cognitive Sciences* **16**:411–417
- Cohen M. A., Dennett D. C., Kanwisher N 2016) **What is the Bandwidth of Perceptual Experience?** *Trends in Cognitive Sciences* **20**:324–335
- Cohen M. A., Keefe J., Brady T. F 2023) **Perceptual awareness occurs along a graded continuum: No evidence of all-or-none failures in continuous reproduction tasks** *Psychological Science* **34**:1033–1047
- De Brigard F., Prinz J. 2010) **Attention and consciousness** *Wiley Interdisciplinary Reviews: Cognitive Science* **1**:51–59
- Dehaene S., Naccache L 2001) **Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework** *Cognition* **79**:1–37
- Dehaene S., Changeux J-P., Naccache L., Sackur J., Sergent S 2006) **Conscious, preconscious, and subliminal processing: a testable taxonomy** *Trends in Cognitive Sciences* **10**:204–211

Devue C., Laloyaux C., Feyers D., Theeuwes J., Brédart S 2009) **Do pictures of faces, and which ones, capture attention in the inattentional-blindness paradigm?** *Perception* **38**:552–568

Drew T., Võ M. L. H., Wolfe J. M 2013) **The invisible gorilla strikes again: Sustained inattentional blindness in expert observers** *Psychological Science* **24**:1848–1853

Dulany D. E 2001) **Inattentional Awareness** *Psyche: An Interdisciplinary Journal of Research on Consciousness* **7**:1–14

Eriksen C.W 1960) **Discrimination and Learning Without Awareness: A Methodological Survey and Evaluation** *Psychological Review* **67**:279–300

Fougnie D., Marois R 2007) **Executive working memory load induces inattentional blindness** *Psychonomic Bulletin & Review* **14**:142–147

Gourevitch V., Galanter E 1967) **A significance test for one parameter isosensitivity functions** *Psychometrika* **32**:25–33

Green D. M., Swets J. A 1966) **Signal detection theory and psychophysics**. New York: Wiley

Hautus M. J 1995) **Corrections for extreme proportions and their biasing effects on estimated values of d'** *Behavior Research Methods, Instruments, & Computers* **27**:46–51

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

Hirschhorn R., Biderman D., Biderman N., Yaron I., Bennet R., Plotnik M., Mudrik L. 2024) **Using virtual reality to induce multi-trial inattentional blindness despite trial-by-trial measures of awareness** *Behavior Research Methods* **56**:3452–3468

Holender D 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

Hollingworth A., Henderson J. M 2002) **Accurate visual memory for previously attended objects in natural scenes** *Journal of Experimental Psychology: Human Perception and Performance* **28**:113–136

Hutchinson B. T 2019) **Toward a theory of consciousness: a review of the neural correlates of inattentional blindness** *Neuroscience & Biobehavioral Reviews* **104**:87–99

Ihaka R., Gentleman R 1996) **R: a language for data analysis and graphics** *Journal of Computational and Graphical Statistics* **5**:299–314

Irvine E. 2012) **Consciousness as a scientific concept: A philosophy of science perspective** Springer Science & Business Media

Jensen M. S., Yao R., Street W. N., Simons D. J 2011) **Change blindness and inattentional blindness** *Wiley Interdisciplinary Reviews: Cognitive Science* **2**:529–546

Kadlec H 1999) **Statistical properties of d' and β estimates of signal detection theory** *Psychological Methods* **4**:22–43

Kolb F. C., Braun J 1995) **Blindsight in normal observers** *Nature* **377**:336–338

- Kreitz C., Pugnaghi G., Memmert D 2020) **Guessing right: Preconscious processing in inattentional blindness** *Quarterly Journal of Experimental Psychology* **73**:1055–1065
- Lamme V. A 2003) **Why visual attention and awareness are different** *Trends in Cognitive Sciences* **7**:12–18
- Levine M. W., McAnany J. J 2005) **The relative capabilities of the upper and lower visual hemifields** *Vision Research* **45**:2820–2830
- Mack A., Rock I 1998) **Inattentional Blindness**. Cambridge, MA: MIT Press
- Macmillan N. A., Creelman C. D 2005) **Detection Theory: A User's Guide**. Cambridge University Press: Cambridge
- Memmert D 2014) **Inattentional blindness to unexpected events in 8–15-year-olds** *Cognitive Development* **32**:103–109
- Mitroff S. R., Simons D. J., Franconeri S. L 2002) **The siren song of implicit change detection** *Journal of Experimental Psychology: Human Perception and Performance* **28**:798–815
- Mitroff S. R., Simons D. J., Levin D. T 2004) **Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information** *Perception & Psychophysics* **66**:1268–1281
- Moore C. M 2001) **Inattentional blindness: Perception or memory and what does it matter** *Psyche* **7**:178–94
- Moore C. M., Egeth H 1997) **Perception without attention: Evidence of grouping under conditions of inattention** *Journal of Experimental Psychology. Human Perception and Performance* **23**:339–352
- Morey R. D., Rouder J. N., Jamil T., Morey M. R. D. 2015) **Package 'BayesFactor'**. <https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf>
- Morgan M. J., Mason A. J. S., Solomon J. A 1997) **Blindsight in normal subjects** *Nature* **385**:401–402
- Most S. B., Scholl B. J., Clifford E. R., Simons D. J 2005) **What you see is what you set: sustained inattentional blindness and the capture of awareness** *Psychological Review* **112**:217–242
- Most S. B., Simons D. J., Scholl B. J., Jimenez R., Clifford E., Chabris C. F 2001) **How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness** *Psychological Science* **12**:9–17
- Murphy G., Greene C. M 2016) **Perceptual load induces inattentional blindness in drivers** *Applied Cognitive Psychology* **30**:479–483
- Murphy H. 2017) **Why We Miss Objects That Are Right in Front of Us** *The New York Times* <https://www.nytimes.com/interactive/2017/10/06/science/why-we-miss-things-in-front-of-us.html>
- Neisser U., Becklen R 1975) **Selective looking: Attending to visually specified events** *Cognitive Psychology* **7**:480–494

- Newell B. R., Shanks D. R (2023) **Open minded: Searching for truth about the unconscious mind.** MIT Press
- Noah S., Mangun G. R (2020) **Recent evidence that attention is necessary, but not sufficient, for conscious perception** *Annals of the New York Academy of Sciences* **1464**:52–63
- Nobre A. d. P., de Melo G. M., Shanks D. R. (2022) **Publication bias casts doubt on implicit processing in inattentional blindness** *Neuroscience and Biobehavioral Reviews* **140**:1–3
- Nobre A. d. P., de Melo G. M., Gauer G., Wagemans J. (2020) **Implicit processing during inattentional blindness: A systematic review and meta-analysis** *Neuroscience and Biobehavioral Reviews* **119**:355–375
- O'Regan J. K., Noë A (2001) **A sensorimotor account of vision and visual consciousness** *Behavioral and Brain Sciences* **24**:939–973
- Pastore R. E., Crawley E. J., Berens M. S., Skelly M. A (2003) **"Nonparametric" A' and other modern misconceptions about signal detection theory** *Psychonomic Bulletin & Review* **10**:556–569
- Peer E., Brandimarte L., Samat S., Acquisti A (2017) **Beyond the Turk: Alternative platforms for crowdsourcing behavioral research** *Journal of Experimental Social Psychology* **70**:153–163
- Phillips I (2021a) **Blindsight is qualitatively degraded conscious vision** *Psychological Review* **128**:558–584
- Phillips I (2021b) **Scepticism about unconscious perception is the default hypothesis** *Journal of Consciousness Studies* **28**:186–205
- Pitts M. A., Martínez A., Hillyard S. A (2012) **Visual processing of contour patterns under conditions of inattentional blindness** *Journal of Cognitive Neuroscience* **24**:287–303
- Pitts M. A., Metzler S., Hillyard S. A (2014) **Isolating neural correlates of conscious perception from neural correlates of reporting one's perception** *Frontiers in Psychology* **5**
- Prinz J (2012) **The Conscious Brain** Oxford University Press
- Prinz J. J., Matthen M. (2015) **Unconscious Perception** *Oxford Handbook of Philosophy of Perception* Oxford: OUP :371–389
- Pugnaghi G., Memmert D., Kreitz C (2020) **Loads of unconscious processing: The role of perceptual load in processing unattended stimuli during inattentional blindness. Attention, Perception & Psychophysics** **82**:2641–2651
- Redlich D., Memmert D., Kreitz C (2021) **A systematic overview of methods, their limitations, and their opportunities to investigate inattentional blindness** *Applied Cognitive Psychology* **35**:136–147
- Rees G., Russell C., Frith C. D., Driver J (1999) **Inattentional blindness versus inattentional amnesia for fixated but ignored words** *Science* **286**:2504–2507
- Rock I., Linnett C. M., Grant P., Mack A (1992) **Perception without attention: Results of a new method** *Cognitive Psychology* **24**:502–534

- Rosenholtz R 2011) **What your visual system sees where you are not looking** *Proceedings of SPIE (Human Vision and Electronic Imaging XVI* **7865**:343–356
- Schnuerch R., Kreitz C., Gibbons H., Memmert D 2016) **Not quite so blind: Semantic processing despite inattention blindness** *Journal of Experimental Psychology: Human Perception and Performance* **42**:459–63
- Scholte H. S., Witteveen S. C., Spekreijse H., Lamme V. A 2006) **The influence of inattention on the neural correlates of scene segmentation** *Brain Research* **1076**:106–115
- Simons D. J 2000) **Attentional capture and inattention blindness** *Trends in Cognitive Sciences* **4**:147–155
- Simons D. J., Chabris C. F 1999) **Gorillas in our midst: Sustained inattention blindness for dynamic events** *Perception* **28**:1059–1074
- Snodgrass M 2002) **Disambiguating conscious and unconscious influences: Do exclusion paradigms demonstrate unconscious perception?** *The American Journal of Psychology* **115**:545–579
- Tanner W. P., Swets J. A. 1954) **A decision-making theory of visual detection** *Psychological Review* **61**:401–409
- Todd J. J., Fougner D., Marois R 2005) **Visual short-term memory load suppresses temporo-parietal junction activity and induces inattention blindness** *Psychological Science* **16**:965–972
- Vandenbroucke A. R., Fahrenfort J. J., Sligte I. G., Lamme V. A 2014) **Seeing without knowing: neural signatures of perceptual inference in the absence of report** *Journal of Cognitive Neuroscience* **26**:955–69
- Ward E. J., Scholl B. J 2015) **Inattention blindness reflects limitations on perception, not memory: Evidence from repeated failures of awareness** *Psychonomic Bulletin & Review* **22**:722–727
- Webb T. W., Graziano M. S. A 2015) **The attention schema theory: a mechanistic account of subjective awareness** *Frontiers in Psychology* **6**:1–11
- Weiskrantz L., Warrington E. K., Sanders M. D., Marshall J 1974) **Visual capacity in the hemianopic field following a restricted occipital ablation** *Brain* **97**:709–728
- White R. C., Davies M., Davies A. M. A 2018) **Inattention blindness on the full-attention trial: Are we throwing out the baby with the bathwater?** *Consciousness & Cognition* **59**:64–77
- Wixted J. T 2020) **The forgotten history of signal detection theory** *Journal of experimental psychology: learning, memory, and cognition* **46**
- Wolfe J. M., Coltheart V. 1999) **Inattentional amnesia** *Fleeting Memories: Cognition of Brief Visual Stimuli* The MIT Press :71–94
- Wolfe J. M., Utochkin I. S 2019) **What is a preattentive feature?** *Current Opinion in Psychology* **29**:19–26

Wood K., Simons D. J 2017) **Selective attention in inattentional blindness: Selection is specific but suppression is not** *Collabra: Psychology* **3**:1–11

Wood K., Simons D. J 2019) **Processing without noticing in inattentional blindness: A replication of Moore and Egeth (1997) and Mack and Rock (1998)** *Attention, Perception & Psychophysics* **81**:1–11

Wu W 2014) **Attention** London and New York: Routledge

Author information

Makaela Nartker

Department of Psychological & Brain Sciences, Johns Hopkins University; Baltimore, United States

ORCID iD: [0000-0003-1817-9048](https://orcid.org/0000-0003-1817-9048)

For correspondence: makaelanartker@gmail.com

Chaz Firestone

Department of Psychological & Brain Sciences, Johns Hopkins University; Baltimore, United States, Department of Philosophy, Johns Hopkins University; Baltimore, United States

ORCID iD: [0000-0002-1247-2422](https://orcid.org/0000-0002-1247-2422)

Howard Egeth

Department of Psychological & Brain Sciences, Johns Hopkins University; Baltimore, United States

ORCID iD: [0000-0002-0147-0827](https://orcid.org/0000-0002-0147-0827)

Ian Phillips

Department of Psychological & Brain Sciences, Johns Hopkins University; Baltimore, United States, Department of Philosophy, Johns Hopkins University; Baltimore, United States

ORCID iD: [0000-0003-2932-8045](https://orcid.org/0000-0003-2932-8045)

For correspondence: ianbphillips@jhu.edu

Editors

Reviewing Editor

Simon van Gaal

University of Amsterdam, Amsterdam, Netherlands

Senior Editor

Timothy Behrens

University of Oxford, Oxford, United Kingdom

Reviewer #1 (Public review):

The results of these experiments support a modest but important conclusion: If sub-optimal methods are used to collect retrospective reports, such as simple yes/no questions, inattentional blindness (IB) rates may be overestimated by up to ~8%.

(1) In experiment 1, data from 374 subjects were included in the analysis. As shown in figure 2b, 267 subjects reported noticing the critical stimulus and 107 subjects reported not noticing it. This translates to a 29% IB rate if we were to only consider the "did you notice anything unusual Y/N" question. As reported in the results text (and figure 2c), when asked to report the location of the critical stimulus (left/right), 63.6% of the "non-noticer" group answered correctly. In other words, 68 subjects were correct about the location while 39 subjects were incorrect. Importantly, because the location judgment was a 2-alternative-forced-choice, the assumption was that if 50% (or at least not statistically different than 50%) of the subjects answered the location question correctly, everyone was purely guessing. Therefore, we can estimate that ~39 of the subjects who answered correctly were simply guessing (because 39 guessed incorrectly), leaving 29 subjects from the non-noticer group who were correct on the 2AFC above and beyond the pure guess rate. If these 29 subjects are moved from the non-noticer to the noticer group, the corrected rate of IB for Experiment 1 is 20.86% instead of the original 28.61% rate that would have been obtained if only the Y/N question was used. In other words, relying only on the "Y/N did you notice anything" question led to an overestimate of IB rates by 7.75% in Experiment 1.

In the revised version of their manuscript, the authors provided the data that was missing from the original submission, which allows this same exercise to be carried out on the other 4 experiments. Using the same logic as above, i.e., calculating the pure-guess rate on the 2AFC, moving the number of subjects above this pure-guess rate to the non-noticer group, and then re-calculating a "corrected IB rate", the other experiments demonstrate the following:

Experiment 2: IB rates were overestimated by 4.74% (original IB rate based only on Y/N question = 27.73%; corrected IB rate that includes the 2AFC = 22.99%)

Experiment 3: IB rates were overestimated by 3.58% (original IB rate = 30.85%; corrected IB rate = 27.27%)

Experiment 4: IB rates were overestimated by ~8.19% (original IB rate = 57.32%; corrected IB rate for color* = 39.71%, corrected IB rate for shape = 52.61%, corrected IB rate for location = 55.07%)

Experiment 5: IB rates were overestimated by ~1.44% (original IB rate = 28.99%; corrected IB rate for color = 27.56%, corrected IB rate for shape = 26.43%, corrected IB rate for location = 28.65%)

*note: the highest overestimate of IB rates was from Experiment 4, color condition, but the authors admitted that there was a problem with 2AFC color guessing bias in this version of the experiment which was a main motivation for running experiment 5 which corrected for this bias.

Taken as a whole, this data clearly demonstrates that even with a conservative approach to analyzing the combination of Y/N and 2AFC data, inattentional blindness was evident in a sizeable portion of the subject populations. An important (albeit modest) overestimate of IB rates was demonstrated by incorporating these improved methods.

(2) One of the strongest pieces of evidence presented in this paper was the single data point in Figure 3e showing that in Experiment 3, even the super subject group that rated their non-noticing as "highly confident" had a d' score significantly above zero. Asking for confidence ratings is certainly an improvement over simple Y/N questions about noticing, and if this result were to hold, it could provide a key challenge to IB. However, this result can most likely be explained by measurement error.

In their revised paper, the authors reported data that was missing from their original submission: the confidence ratings on the 2AFC judgments that followed the initial Y/N

question. The most striking indication that this data is likely due to measurement error comes from the number of subjects who indicated that they were highly confident that they didn't notice anything on the critical trial, but then when asked to guess the location of the stimulus, indicated that they were highly confident that the stimulus was on the left (or right). There were 18 subjects (8.82% of the high-confidence non-noticer group) who responded this way. To most readers, this combination of responses (high confidence in correctly judging a stimulus feature that one is highly confident in having not seen at all) indicates that a portion of subjects misunderstood the confidence scales (or just didn't read the questions carefully or made mistakes in their responses, which is common for experiments conducted online).

In the authors' rebuttal to the first round of peer review, they wrote, "it is perfectly rationally coherent to be very confident that one didn't see anything but also very confident that if there was anything to be seen, it was on the left." I respectfully disagree that such a combination of responses is rationally coherent. The more parsimonious interpretation is that a measurement error occurred, and it's questionable whether we should trust any responses from these 18 subjects.

In their rebuttal, the authors go on to note that 14 of the 18 subjects who rated their 2AFC with high confidence were correct in their location judgment. If these 14 subjects were removed from analysis (which seems like a reasonable analysis choice, given their contradictory responses), d' for the high-confidence non-noticer group would most likely fall to chance levels. In other words, we would see a data pattern similar to that plotted in Figure 3e, but with the first data point on the left moving down to zero d' . This corrected Figure 3e would then provide a very nice evidence-based justification for including confidence ratings along with Y/N questions in future inattentional blindness studies.

(3) In most (if not all) IB experiments in the literature, a partial attention and/or full attention trial is administered after the critical trial. These control trials are very important for validating IB on the critical trial, as they must show that, when attended, the critical stimuli are very easy to see. If a subject cannot detect the critical stimulus on the control trial, one cannot conclude that they were inattentionally blind on the critical trial, e.g., perhaps the stimulus was just too difficult to see (e.g., too weak, too brief, too far in the periphery, too crowded by distractor stimuli, etc.), or perhaps they weren't paying enough attention overall or failed to follow instructions. In the aggregate data, rates of noticing the stimuli should increase substantially from the critical trial to the control trials. If noticing rates are equivalent on the critical and control trials, one cannot conclude that attention was manipulated in the first place.

In their rebuttal to the first round of peer review, the authors provided weak justification for not including such a control condition. They cite one paper that argues such control conditions are often used to exclude subjects from analysis (those who fail to notice the stimulus on the control trial are either removed from analysis or replaced with new subjects) and such exclusions/replacements can lead to underestimations of inattentional blindness rates. However, the inclusion of a partial or full attention condition as a control does not necessitate the extra step of excluding or replacing subjects. In the broadest sense, such a control condition simply validates the attention manipulation, i.e., one can easily compare the percent of subjects who answered "yes" or who got the 2AFC judgment correct during the critical trial versus the control trial. The subsequent choice about exclusion/replacement is separate, and researchers can always report the data with and without such exclusions/replacements to remain more neutral on this practice.

If anyone were to follow-up on this study, I highly recommend including a partial or full attention control condition, especially given the online nature of data collection. It's important to know the percent of online subjects who answer yes and who get the 2AFC question correct when the critical stimulus is attended, because that is the baseline (in this

case, the "ceiling level" of performance) to which the IB rates on the critical trial can be compared.

<https://doi.org/10.7554/eLife.100337.2.sa2>

Reviewer #2 (Public review):

In this study, Nartker et al. examine how much observers are conscious of using variations of classic inattentional blindness studies. The key idea is that rather than simply ask observers if they noticed a critical object with one yes/no question, the authors also ask follow-up questions to determine if observers are aware of more than the yes/no questions suggest. Specifically, by having observers make forced choice guesses about the critical object, the authors find that many observers who initially said "no" they did not see the object can still "guess" above chance about the critical object's location, color, etc. Thus, the authors claim, that prior claims of inattentional blindness are mistaken and that using such simple methods has led numerous researchers to overestimate how little observers see in the world. To quote the authors themselves, these results imply that "inattentionally blind subjects consciously perceive these stimuli after all... they show sensitivity to IB stimuli because they can see them."

Before getting to a few issues I have with the paper, I do want to make sure to explicitly compliment the researchers for many aspects of their work. Getting massive amounts of data, using signal detection measures, and the novel use of a "super subject" are all important contributions to the literature that I hope are employed more in the future.

Main point 1: My primary issue with this work is that I believe the authors are misrepresenting the way people often perform inattentional blindness studies. In effect, the authors are saying, "People do the studies 'incorrectly' and report that people see very little. We perform the studies 'correctly' and report that people see much more than previously thought." But the way previous studies are conducted is not accurately described in this paper. The authors describe previous studies as follows on page 3:

"Crucially, however, this interpretation of IB and the many implications that follow from it rest on a measure that psychophysics has long recognized to be problematic: simply asking participants whether they noticed anything unusual. In IB studies, awareness of the unexpected stimulus (the novel shape, the parading gorilla, etc.) is retroactively probed with a yes/no question, standardly, "Did you notice anything unusual on the last trial which wasn't there on previous trials?". Any subject who answers "no" is assumed not to have any awareness of the unexpected stimulus.

If this quote were true, the authors would have a point. Unfortunately, I do not believe it is true. This is simply not how many inattentional blindness studies are run. Some of the most famous studies in the inattentional blindness literature do not simply as observes a yes/no question (e.g., the invisible gorilla (Simons et al. 1999), the classic door study where the person changes (Simons and Levin, 1998), the study where observers do not notice a fight happening a few feet from them (Chabris et al., 2011). Instead, these papers consistently ask a series of follow-up questions and even tell the observers what just occurred to confirm that observers did not notice that critical event (e.g., "If I were to tell you we just did XYZ, did you notice that?"). In fact, after a brief search on Google Scholar, I was able to relatively quickly find over a dozen papers that do not just use a yes/no procedure, and instead as a series of multiple questions to determine if someone is inattentionally blind. In no particular order some papers:

- (1) Most et al. (2005) Psych Review
- (2) Drew et al. (2013) Psych Science

- (3) Drew et al. (2016) Journal of Vision
- (4) Simons et al. (1999) Perception
- (5) Simons and Levin (1998) Perception
- (6) Chabris et al. (2011) iPerception
- (7) Ward & Scholl (2015) Psych Bulletin and Review
- (8) Most et al. (2001) Psych Science
- (9) Todd & Marois (2005) Psych Science
- (10) Fougner & Marois (2007) Psych Bulletin and Review
- (11) New and German (2015) Evolution and Human Behaviour
- (12) Jackson-Nielsen (2017) Consciousness and cognition
- (13) Mack et al. (2016) Consciousness and cognition
- (14) Devue et al. (2009) Perception
- (15) Memmert (2014) Cognitive Development
- (16) Moore & Egeth (1997) JEP:HPP
- (17) Cohen et al. (2020) Proc Natl Acad Sci
- (18) Cohen et al. (2011) Psych Science

This is a critical point. The authors' key idea is that when you ask more than just a simple yes/no question, you find that other studies have overestimated the effects of inattention blindness. But none of the studies listed above only asked simple yes/no questions. Thus, I believe the authors are mis-representing the field. Moreover, many of the studies that do much more than ask a simple yes/no question are cited by the authors themselves! Furthermore, as far as I can tell, the authors believe that if researchers do these extra steps and ask more follow-ups, then the results are valid. But since so many of these prior studies do those extra steps, I am not exactly sure what is being criticized.

To make sure this point is clear, I'd like to use a paper of mine as an example. In this study (Cohen et al., 2020, Proc Natl Acad Sci USA) we used gaze-contingent virtual reality to examine how much color people see in the world. On the critical trial, the part of the scene they fixated on was in color, but the periphery was entirely in black and white. As soon as the trial ended, we asked participants a series of questions to determine what they noticed. The list of questions included:

- (1) "Did you notice anything strange or different about that last trial?"
- (2) "If I were to tell you that we did something odd on the last trial, would you have a guess as to what we did?"
- (3) "If I were to tell you we did something different in the second half of the last trial, would you have a guess as to what we did?"
- (4) "Did you notice anything different about the colors in the last scene?"
- (5) We then showed observers the previous trial again and drew their attention to the effect and confirmed that they did not notice that previously.

In a situation like this, when the observers are asked so many questions, do the authors believe that "the inattentionally blind can see after all?" I believe they would not say that and the reason they would not say that is because of the follow-up questions after the initial yes/no question. But since so many previous studies use similar follow-up questions, I do not think you can state that the field is broadly overestimating inattention blindness. This is why it seems to me to be a bit of a straw-man: most people do not just use the yes/no method.

Main point 2: Let's imagine for a second that every study did just ask a yes/no question and then would stop. So, the criticism the authors are bringing up is valid (even though I believe it is not). I am not entirely sure that above chance performance on a forced choice task proves that the inattentionally blind can see after all. Could it just be a form of subliminal priming? Could there be a significant number of participants who basically would say something like, "No I did not see anything, and I feel like I am just guessing, but if you want me to say whether the thing was to the left or right, I will just 100% guess"? I know the literature on

priming from things like change and inattentional blindness is a bit unclear, but this seems like maybe what is going on. In fact, maybe the authors are getting some of the best priming from inattentional blindness because of their large sample size, which previous studies do not use.

I'm curious how the authors would relate their studies to masked priming. In masked priming studies, observers say they did not see the target (like in this study) but still are above chance when forced to guess (like in this study). Do the researchers here think that that is evidence of "masked stimuli are truly seen" even if a participant openly says they are guessing?

Main point 3: My last question is about how the authors interpret a variety of inattentional blindness findings. Previous work has found that observers fail to notice a gorilla in a CT scan (Drew et al., 2013), a fight occurring right in front of them (Chabris et al., 2011), a plane on a runway that pilots crash into (Haines, 1991), and so forth. In a situation like this, do the authors believe that many participants are truly aware of these items but simply failed to answer a yes/no question correctly? For example, imagine the researchers made participants choose if the gorilla was in the left or right lung and some participants who initially said they did not notice the gorilla were still able to correctly say if it was in the left or right lung. Would the authors claim "that participant actually did see the gorilla in the lung"? I ask because it is difficult to understand what it means to be aware of something as salient as a gorilla in a CT scan, but say "no" you didn't notice it when asked a yes/no question. What does it mean to be aware of such important, ecologically relevant stimuli, but not act in response to them and openly say "no" you did not notice them?

Overall: I believe there are many aspects of this set of studies that are innovative and I hope the methods will be used more broadly in the literature. However, I believe the authors misrepresent the field and overstate what can be interpreted from their results. While I am sure there are cases where more nuanced questions might reveal inattentional blindness is somewhat overestimated, claims like "the inattentionally blind can see after all" or "Inattentionally blind subjects consciously perceive the stimuli after all" seem to be incorrect (or at least not at all proven by this data).

<https://doi.org/10.7554/eLife.100337.2.sa1>

Author response:

The following is the authors' response to the current reviews.

Responses to Reviewer #1:

We thank the reviewer for these additional comments, and more generally for their extensive engagement with our work, which is greatly appreciated. Here, we respond to the three points in their latest review in turn.

The results of these experiments support a modest but important conclusion: If sub-optimal methods are used to collect retrospective reports, such as simple yes/no questions, inattentional blindness (IB) rates may be overestimated by up to ~8%.

It is true, of course, that we think the field has overstated the extent of IB, and we appreciate the reviewer characterizing our results as important along these lines. Nevertheless, we respectfully disagree with the framing and interpretation the reviewer attaches to them. As explained in our previous response, we think this interpretation — and the associated calculations of IB overestimation 'rates' — perpetuates a binary approach to perception and awareness which we regard as mistaken.

A graded approach to IB and visual awareness

Our sense is that many theorists interested in IB have conceived of perception and awareness as ‘all or nothing’: You either see a perfectly clear gorilla right in front of you, or you see nothing at all. This is implicit in the reviewer’s characterization of our results as simply indicating that fewer subjects fail to see the critical stimulus than previously assumed. To think that way is precisely to assume the orthodox binary position about perception, i.e., that any given subject can neatly be categorized into one of two boxes, *saw* or *didn’t see*.

Our perspective is different. We think there can be degraded forms of perception and awareness that fall neatly into neither of the categories “saw the stimulus perfectly clearly” or “saw nothing at all”. On this graded conception, the question is not: “What proportion of subjects saw the stimulus?” but: “What is the sensitivity of subjects to the stimulus?” This is why we prefer signal detection measures like d' over % noticing and % correct. This powerful framework has been successful in essentially every domain to which it has been applied, and we think perception and visual awareness are no exception. We understand that the reviewer may not think the same way about this foundational issue, but since part of our goal is to promote a graded approach to perception, we are keen to highlight our disagreement here and so resist the reviewer’s interpretation of our results (even to the extent that it is a positive one!).

Finally, we note that given this perspective, we are correspondingly inclined to reject many of the summary figures following below in Point (1) by the reviewer. These calculations (given in terms of % noticing and not noticing) make sense on the binary conception of awareness, but not on the SDT-based approach we favor. We say more about this below.

(1) In experiment 1, data from 374 subjects were included in the analysis. As shown in figure 2b, 267 subjects reported noticing the critical stimulus and 107 subjects reported not noticing it. This translates to a 29% IB rate if we were to only consider the "did you notice anything unusual Y/N" question. As reported in the results text (and figure 2c), when asked to report the location of the critical stimulus (left/right), 63.6% of the "non-noticer" group answered correctly. In other words, 68 subjects were correct about the location while 39 subjects were incorrect. Importantly, because the location judgment was a 2-alternative-forced-choice, the assumption was that if 50% (or at least not statistically different than 50%) of the subjects answered the location question correctly, everyone was purely guessing. Therefore, we can estimate that ~39 of the subjects who answered correctly were simply guessing (because 39 guessed incorrectly), leaving 29 subjects from the nonnoticer group who were correct on the 2AFC above and beyond the pure guess rate. If these 29 subjects are moved from the non-noticer to the noticer group, the corrected rate of IB for Experiment 1 is 20.86% instead of the original 28.61% rate that would have been obtained if only the Y/N question was used. In other words, relying only on the "Y/N did you notice anything" question led to an overestimate of IB rates by 7.75% in Experiment 1.

In the revised version of their manuscript, the authors provided the data that was missing from the original submission, which allows this same exercise to be carried out on the other 4 experiments.

(To briefly interject: All of these data were provided in our public archive since our original submission and remain available at <https://osf.io/fcrhu>. The difference now is only that they are included in the manuscript itself.)

Using the same logic as above, i.e., calculating the pure-guess rate on the 2AFC, moving the number of subjects above this pure-guess rate to the non-noticer group, and then re-calculating a "corrected IB rate", the other experiments demonstrate the following:

Experiment 2: IB rates were overestimated by 4.74% (original IB rate based only on Y/N question = 27.73%; corrected IB rate that includes the 2AFC = 22.99%)

Experiment 3: IB rates were overestimated by 3.58% (original IB rate = 30.85%; corrected IB rate = 27.27%)

Experiment 4: IB rates were overestimated by ~8.19% (original IB rate = 57.32%; corrected IB rate for color = 39.71%, corrected IB rate for shape = 52.61%, corrected IB rate for location = 55.07%)*

Experiment 5: IB rates were overestimated by ~1.44% (original IB rate = 28.99%; corrected IB rate for color = 27.56%, corrected IB rate for shape = 26.43%, corrected IB rate for location = 28.65%)

**note: the highest overestimate of IB rates was from Experiment 4, color condition, but the authors admitted that there was a problem with 2AFC color guessing bias in this version of the experiment which was a main motivation for running experiment 5 which corrected for this bias.*

Taken as a whole, this data clearly demonstrates that even with a conservative approach to analyzing the combination of Y/N and 2AFC data, inattentive blindness was evident in a sizeable portion of the subject populations. An important (albeit modest) overestimate of IB rates was demonstrated by incorporating these improved methods.

We appreciate the work the reviewer has put into making these calculations. However, as noted above, such calculations implicitly reflect the binary approach to perception and awareness that we reject.

Consider how we'd think about the single subject case where the task is 2afc detection of a low contrast stimulus in noise. Suppose that this subject achieves 70% correct. One way of thinking about this is that the subject fully and clearly sees the stimulus on 40% of trials (achieving 100% correct on those) and guesses completely blindly on the other 60% (achieving 50% correct on those) for a total of $40\% + 30\% = 70\%$ overall. However, this is essentially a 'high threshold' approach to the problem, in contrast to an SDT approach. On an SDT approach — an approach with tremendous evidential support — on every trial the subject receives samples from probabilistic distributions corresponding to each interval (one noise and one signal + noise) and determines which is higher according to the 2afc decision rule. Thus, across trials, they have access to differentially graded information about the stimulus. Moreover, on some trials they may have significant information from the stimulus (perhaps, well above their single interval detection criterion) but still decide incorrectly because of high noise from the other spatial interval. From this perspective, there is no nonarbitrary way of saying whether the subject saw/did not see on a given trial. Instead, we must characterize the subject's overall sensitivity to the stimulus/its visibility to them in terms of a parameter such as d' (here, ~ 0.7).

We take the same attitude to the subjects in our experiments (and specifically to our 'super subject'). Instead of calculating the proportion of subjects who saw or failed to see the stimulus (with some characterized as aware and some as unaware), we think the best way to characterize our results is that, across subjects (and so trials also), there was differential graded access to information from the stimulus, and this is best represented in terms of the group-level sensitivity parameter d' . This is why we frame our results as demonstrating that subjects traditionally considered inattentionally blind exhibit significant residual visual sensitivity to the critical stimulus.

(2) One of the strongest pieces of evidence presented in this paper was the single data point in Figure 3e showing that in Experiment 3, even the super subject group that rated

their non-noticing as "highly confident" had a d' score significantly above zero. Asking for confidence ratings is certainly an improvement over simple Y/N questions about noticing, and if this result were to hold, it could provide a key challenge to IB. However, this result can most likely be explained by measurement error.

In their revised paper, the authors reported data that was missing from their original submission: the confidence ratings on the 2AFC judgments that followed the initial Y/N question. The most striking indication that this data is likely due to measurement error comes from the number of subjects who indicated that they were highly confident that they didn't notice anything on the critical trial, but then when asked to guess the location of the stimulus, indicated that they were highly confident that the stimulus was on the left (or right). There were 18 subjects (8.82% of the high-confidence non-noticer group) who responded this way. To most readers, this combination of responses (high confidence in correctly judging a stimulus feature that one is highly confident in having not seen at all) indicates that a portion of subjects misunderstood the confidence scales (or just didn't read the questions carefully or made mistakes in their responses, which is common for experiments conducted online).

In the authors' rebuttal to the first round of peer review, they wrote, "it is perfectly rationally coherent to be very confident that one didn't see anything but also very confident that if there was anything to be seen, it was on the left." I respectfully disagree that such a combination of responses is rationally coherent. The more parsimonious interpretation is that a measurement error occurred, and it's questionable whether we should trust any responses from these 18 subjects.

In their rebuttal, the authors go on to note that 14 of the 18 subjects who rated their 2AFC with high confidence were correct in their location judgment. If these 14 subjects were removed from analysis (which seems like a reasonable analysis choice, given their contradictory responses), d' for the high-confidence non-noticer group would most likely fall to chance levels. In other words, we would see a data pattern similar to that plotted in Figure 3e, but with the first data point on the left moving down to zero d' . This corrected Figure 3e would then provide a very nice evidence-based justification for including confidence ratings along with Y/N questions in future inattentional blindness studies.

We appreciate the reviewer's highlighting of this particular piece of evidence as amongst our strongest. (At the same time, we must resist its characterization as a "single data point": it derives from a large pre-registered experiment involving some 7,000 subjects total, with over 200 subjects in the relevant bin — both figures being far larger than a typical IB experiment.) We also appreciate their raising the issue of measurement error.

Specifically, the reviewer contends that our finding that even highly confident non-noticers exhibit significant sensitivity is "most likely ... explained by measurement error" due to subjects mistakenly inverting our confidence scale in giving their response. In our original reply, we gave two reasons for thinking this quite unlikely; the reviewer has not addressed these in this revised review. First, we explicitly labeled our confidence scale (with 0 labeled as 'Not at all confident' and 3 as 'Highly confident') so that subjects would be very unlikely simply to invert the scale. This is especially so as it is very counterintuitive to treat "0" as reflecting *high* confidence. More importantly, however, we reasoned that any measurement error due to inverting or misconstruing the confidence scale should be symmetric. That is: If subjects are liable to invert the confidence scale, they should do so just as often when they answer "yes" as when they answer "no" — after all the very same scale is being used in both cases. This allows us to explore evidence of measurement error in relation to the large number of high-confidence "yes" subjects ($N = 2677$), thus providing a robust indicator as to whether subjects are generally liable to misconstrue the confidence scale. Looking at the

number of such high confidence noticers who subsequently respond to the 2afc question with low confidence (a pattern which might, though need not, suggest measurement error), we found that the number was tiny. Only 28/2677 (1.05%) of high-confidence noticers subsequently gave the lowest level of confidence on the 2afc question, and only 63/2677 (2.35%) subjects gave *either* of the two lower levels of confidence. For these reasons, we consider any measurement error due to misunderstanding the confidence scale to be extremely minimal.

The reviewer is correct to note that 18/204 (9%) subjects reported both being highly confident that they didn't notice anything and highly confident in their 2afc judgment, although only 14/18 were correct in this judgment. Should we exclude these 14? Perhaps if we agree with the reviewer that such a pattern of responses is not "rationally coherent" and so must reflect a misconstrual of the scale. But such a pattern is in fact perfectly and straightforwardly intelligible. Specifically, in a 2afc task, two stimuli can individually fall well below a subject's single interval detection criterion — leading to a high confidence judgment that nothing was presented in either interval. Quite consistent with this, the lefthand stimulus may produce a signal that is much higher than the right-hand stimulus — leading to a high confidence forced-choice judgment that, *if something was presented*, it was on the left. (By analogy, consider how a radiologist could look at a scan and say the following: "We're 95% confident there's no tumor. But even on the 5% chance that there is, our tests completely rule out that it's a malignant one, so don't worry.")

(3) In most (if not all) IB experiments in the literature, a partial attention and/or full attention trial is administered after the critical trial. These control trials are very important for validating IB on the critical trial, as they must show that, when attended, the critical stimuli are very easy to see. If a subject cannot detect the critical stimulus on the control trial, one cannot conclude that they were inattentionally blind on the critical trial, e.g., perhaps the stimulus was just too difficult to see (e.g., too weak, too brief, too far in the periphery, too crowded by distractor stimuli, etc.), or perhaps they weren't paying enough attention overall or failed to follow instructions. In the aggregate data, rates of noticing the stimuli should increase substantially from the critical trial to the control trials. If noticing rates are equivalent on the critical and control trials, one cannot conclude that attention was manipulated in the first place.

In their rebuttal to the first round of peer review, the authors provided weak justification for not including such a control condition. They cite one paper that argues such control conditions are often used to exclude subjects from analysis (those who fail to notice the stimulus on the control trial are either removed from analysis or replaced with new subjects) and such exclusions/replacements can lead to underestimations of inattentional blindness rates. However, the inclusion of a partial or full attention condition as a control does not necessitate the extra step of excluding or replacing subjects. In the broadest sense, such a control condition simply validates the attention manipulation, i.e., one can easily compare the percent of subjects who answered "yes" or who got the 2AFC judgment correct during the critical trial versus the control trial. The subsequent choice about exclusion/replacement is separate, and researchers can always report the data with and without such exclusions/replacements to remain more neutral on this practice.

If anyone were to follow-up on this study, I highly recommend including a partial or full attention control condition, especially given the online nature of data collection. It's important to know the percent of online subjects who answer yes and who get the 2AFC question correct when the critical stimulus is attended, because that is the baseline (in this case, the "ceiling level" of performance) to which the IB rates on the critical trial can be compared.

We agree with the reviewer that future studies could benefit from including a partial or full attention condition. They are surely right that we might learn something additional from such conditions.

Where we differ from the reviewer is in thinking of these conditions as “controls” appropriate to our research question. This is why we offered the justification we did in our earlier response. *When these conditions are used as controls*, they are used to exclude subjects in ways that serve to inflate the biases we are concerned with in our work. For our question, the absence of these conditions does not impact the significance of the findings, since such conditions are designed to answer a question which is not the one at the heart of our paper. Our key claim is that subjects who deny noticing an unexpected stimulus in a standard inattentional blindness paradigm nonetheless exhibit significant residual sensitivity (as well as a conservative bias in their response to the noticing question); the presence or absence of partial- or full-attention conditions is orthogonal to that question.

Moreover, we note that our tasks were precisely chosen to be classic tasks widely used in the literature to manipulate attention. Thus, by common consensus in the field, they are effective means to soak up attention, and have in effect been tested in partial- and full-attention control settings in a huge number of studies. Second, we think it very doubtful that subjects in a full-attention trial would not overwhelmingly have detected our critical stimuli. The reviewer worries that they might have been “too weak, too brief, too far in the periphery, too crowded by distractor stimuli, etc.” But consider E5 where the stimulus was a highly salient orange or green shape, present on the screen for 5 seconds. The reviewer also suggests that subjects in the full-attention control might not have detected the stimulus because they “weren’t paying enough attention overall”. But evidently if they weren’t paying attention even in the full-attention trial this would be reason for thinking that there was inattentional blindness even in this condition (a point made by White et al. 2018) and certainly not a reason for thinking there was not an attentional effect in the critical trial. Lastly, the reviewer suggests that a full-attention condition would have helped ensure that subjects were following instructions. But we ensured this already by (as per our pre-registration) excluding subjects who performed poorly in the relevant primary tasks.

Thus, both in principle and in practice, we do not see the absence of such conditions as impacting the interpretation of our findings, even as we agree that future work posing a different research question could certainly learn something from including such conditions.

Responses to Reviewer #2:

We note that this report is unchanged from an earlier round of review, and not a response to our significantly revised manuscript. We believe our latest version fully addresses all the issues which the reviewer originally raised. The interested reader can see our original response below. We again thank the reviewer for their previous report which was extremely helpful.

The following is the authors’ response to the original reviews.

eLife Assessment

This study presents valuable findings to the field interested in inattentional blindness (IB), reporting that participants indicating no awareness of unexpected stimuli through yes/no questions, still show above-chance sensitivity to specific properties of these stimuli through follow-up forced-choice questions (e.g., its color). The results suggest that this is because participants are conservative and biased to report not noticing in IB. The authors conclude that these results provide evidence for residual perceptual awareness of inattentionally blind stimuli and that therefore these findings cast doubt on the claim that awareness requires attention. Although the samples are large and the analysis

protocol novel, the evidence supporting this interpretation is still incomplete, because effect sizes are rather small, the experimental design could be improved and alternative explanations have not been ruled out.

We are encouraged to hear that *eLife* found our work “valuable”. We also understand, having closely looked at the reviews, why the assessment also includes an evaluation of “incomplete”. We gave considerable attention to this latter aspect of the assessment in our revision. In addition to providing additional data and analyses that we believe strengthen our case, we also include a much more substantial review and critique of existing methods in the IB literature to make clear exactly the gap our work fills and the advance it makes. (Indeed, if it is appropriate to say this here, we believe one key aspect of our work that is missing from the assessment is our inclusion of ‘absent’ trials, which is what allows us to make the crucial claims about conservative reporting of awareness in IB for the first time.) Moreover, we refocus our discussion on only our most central claims, and weaken several of our secondary claims so that the data we’ve collected are better aligned with the conclusions we draw, to ensure that the case we now make is in fact complete. Specifically, our two core claims are (1) that there is residual sensitivity to visual features for subjects who would ordinarily be classified as inattentionally blind (whether this sensitivity is conscious or not), and (2) that there is a tendency to respond conservatively on yes/no questions in the context of IB. We believe we have very compelling support for these two core claims, as we explain in detail below and also through revisions to our manuscript.

Given the combination of strengthened and clarified case, as well as the weakening of any conclusions that may not have been fully supported, we believe and hope that these efforts make our contribution “solid”, “convincing”, or even “compelling” (especially because the “compelling” assessment characterizes contributions that are “more rigorous than the current state-of-the-art”, which we believe to be the case given the issues that have plagued this literature and that we make progress on).

Reviewer #1 (Public review):

Summary:

In the abstract and throughout the paper, the authors boldly claim that their evidence, from the largest set of data ever collected on inattentional blindness, supports the views that "inattentionally blind participants can successfully report the location, color, and shape of stimuli they deny noticing", "subjects retain awareness of stimuli they fail to report", and "these data...cast doubt on claims that awareness requires attention." If their results were to support these claims, this study would overturn 25+ years of research on inattentional blindness, resolve the rich vs. sparse debate in consciousness research, and critically challenge the current majority view in cognitive science that attention is necessary for awareness.

Unfortunately, these extraordinary claims are not supported by extraordinary (or even moderately convincing) evidence. At best, the results support the more modest conclusion: If sub-optimal methods are used to collect retrospective reports, inattentional blindness rates will be overestimated by up to ~8% (details provided below in comment #1). This evidence-based conclusion means that the phenomenon of inattentional blindness is alive and well as it is even robust to experiments that were specifically aimed at falsifying it. Thankfully, improved methods already exist for correcting the ~8% overestimation of IB rates that this study successfully identified.

We appreciate here the reviewer’s recognition of the importance of work on inattentional blindness, and the centrality of inattentional blindness to a range of major questions. We also recognize their concerns with what they see as a gap between our data and the claims made

on their basis. We address this in detail below (as well as, of course, in our revised manuscript). However, from the outset we are keen to clarify that our central claim is only the first one the reviewer mentions — and the one which appears in our title — namely that, as a group, participants can successfully report the location, color, and shape of stimuli they deny noticing, and thus that there is “Sensitivity to visual features in inattentional blindness”. This is the claim that we believe is strongly supported by our data, and all the more so after revising the manuscript in light of the helpful comments we’ve received.

By contrast, the other claims the reviewer mentions, concerning awareness (as opposed to residual sensitivity—which might be conscious or unconscious) were intended as both secondary and tentative. We agree with the referee that these are not as strongly supported by our data (and indeed we say so in our manuscript), whereas we do think our data strongly support the more modest — and, to us central — claim that, as a group, inattentionally blind participants can successfully report the location, color, and shape of stimuli they deny noticing.

We also feel compelled to resist somewhat the reviewer’s summary of our claims. For example, the reviewer attributes to us the claim that “subjects retain awareness of stimuli they fail to report”; but while that phrase does appear in our abstract, what we in fact say is that our data are “**consistent with** an alternative hypothesis about IB, namely that subjects retain awareness of stimuli they fail to report”. We do in fact believe that our data are consistent with that hypothesis, whereas earlier investigations seemed not to be. We mention this only because we had used that careful phrasing precisely for this sort of reason, so that we *wouldn’t* be read as saying that our results unequivocally support that alternative.

Still, looking back, we see how we may have given more emphasis than we intended to some of these more secondary claims. So, we’ve now gone through and revised our manuscript throughout to emphasize that our main claim is about residual sensitivity, and to make clear that our claims about awareness are secondary and tentative. Indeed, we now say precisely this, that although we favor an interpretation of “our results in terms of residual conscious vision in IB ... this claim is tentative and secondary to our primary finding”. We also weaken the statements in the abstract that the reviewer mentions, to better reflect our key claims.

Finally, we note one further point: Dialectically, inattentional blindness has been used to argue (e.g.) that attention is required for awareness. We think that our data concerning residual sensitivity at least push back on the *use* of IB to make this claim, even if (as we agree) they do not provide decisive evidence that awareness survives inattention. In other words, we think our data call that claim into question, such that it’s now genuinely unclear whether awareness does or does not survive inattention. We have adjusted our claims on this point accordingly as well.

Comments:

(1) In experiment 1, data from 374 subjects were included in the analysis. As shown in figure 2b, 267 subjects reported noticing the critical stimulus and 107 subjects reported not noticing it. This translates to a 29% IB rate, if we were to only consider the "did you notice anything unusual Y/N" question. As reported in the results text (and figure 2c), when asked to report the location of the critical stimulus (left/right), 63.6% of the "non-noticer" group answered correctly. In other words, 68 subjects were correct about the location while 39 subjects were incorrect. Importantly, because the location judgment was a 2-alternative-forced-choice, the assumption was that if 50% (or at least not statistically different than 50%) of the subjects answered the location question correctly, everyone was purely guessing. Therefore, we can estimate that ~39 of the subjects who answered correctly were simply guessing (because 39 guessed incorrectly), leaving 29 subjects from the nonnoticer group who may have indeed actually seen the location of the stimulus. If these 29 subjects are moved to the noticer group, the corrected rate of IB

for experiment 1 is 21% instead of 29%. In other words, relying only on the "Y/N did you notice anything" question leads to an overestimate of IB rates by 8%. This modest level of inaccuracy in estimating IB rates is insufficient for concluding that "subjects retain awareness of stimuli they fail to report", i.e. that inattentional blindness does not exist.

In addition, this 8% inaccuracy in IB rates only considers one side of the story. Given the data reported for experiment 1, one can also calculate the number of subjects who answered "yes, I did notice something unusual" but then reported the incorrect location of the critical stimulus. This turned out to be 8 subjects (or 3% of the "noticer" group). Some would argue that it's reasonable to consider these subjects as inattentionally blind, since they couldn't even report where the critical stimulus they apparently noticed was located. If we move these 8 subjects to the non-noticer group, the 8% overestimation of IB rates is reduced to 6%.

The same exercise can and should be carried out on the other 4 experiments, however, the authors do not report the subject numbers for any of the other experiments, i.e., how many subjects answered Y/N to the noticing question and how many in each group correctly answered the stimulus feature question. From the limited data reported (only total subject numbers and d' values), the effect sizes in experiments 2-5 were all smaller than in experiment 1 (d' for the non-noticer group was lower in all of these follow-up experiments), so it can be safely assumed that the ~6-8% overestimation of IB rates was smaller in these other four experiments. In a revision, the authors should consider reporting these subject numbers for all 5 experiments.

We now report, as requested, all these subject numbers in our supplementary data (see Supplementary Tables 1 and 2 in our Supplementary Materials).

However, we wish to address the larger question the reviewer has raised: Do our data only support a relatively modest reduction in IB rates? Even if they did, we *still* believe that this would be a consequential result, suggesting a significant overestimation of IB rates in classic paradigms. However, part of our purpose in writing this paper is to push back against a certain binary way of thinking about seeing/awareness. Our sense is that the field has conceived of awareness as "all or nothing": You either see a perfectly clear gorilla right in front of you, or you see nothing at all. Our perspective is different: We think there can be degraded forms of awareness that fall into neither of those categories. For that reason, we are disinclined to see our results in the way that the reviewer suggests, namely as simply indicating that fewer subjects fail to see the stimulus than previously assumed. To think that way is, in our view, to assume the orthodox binary position about awareness. If, instead, one conceives of awareness as we do (and as we believe the framework of signal detection theory should compel us to), then it isn't quite right to think of the proportion of subjects who were aware, but rather (e.g.) the sensitivity of subjects to the relevant stimulus. This is why we prefer measures like d' over % noticing and % correct. We understand that the reviewer may not think the same way about this issue as we do, but part of our goal is to promote that way of thinking in general, and so some of our comments below reflect that perspective and approach.

For example, consider how we'd think about the single subject case where the task is 2afc detection of a low contrast stimulus in noise. Suppose that this subject achieves 70% correct. One way of thinking about that is that the subject sees the stimulus on 40% of trials (achieving 100% correct on those) and guesses blindly on the other 60% (achieving 50% correct on those) for a total of $40\% + 30\% = 70\%$ overall. However, this is essentially a "high threshold" approach to the problem, in contrast to an SDT approach. On an SDT approach (an approach with tremendous evidential support), on every trial the subject receives samples from probabilistic distributions corresponding to each interval (one noise and one signal + noise) and determines which is higher according to the 2afc decision rule. Thus, across trials

they have access to differentially graded information about the stimulus. Moreover, on some trials they may have significant information from the stimulus (perhaps, well above their single interval detection criterion) but still decide incorrectly because of high noise from the other spatial interval. From this perspective, there is no non-arbitrary way of saying whether the subject saw/did not see on a given trial. Instead, we must characterize the subject's overall sensitivity to the stimulus/its visibility to them in terms of a parameter such as d' (here, ~ 0.7).

We take the same attitude to our super subject. Instead of saying that some subjects saw/failed to see the stimuli, instead we suggest that the best way to characterize our results is that across subjects (and so trials also) there was differential graded access to information from the stimulus best represented in terms of the group-level sensitivity parameter d' .

We acknowledge that (despite ourselves) we occasionally fell into an all-too-natural binary/high threshold way of thinking, as when we suggested that our data show that "inattentionally blind subjects consciously perceive these stimuli after all" and "the inattentionally blind can see after all." (p.17) We have removed such problematic phrasing as well as other problematic phrasing as noted below.

(2) Because classic IB paradigms involve only one critical trial per subject, the authors used a "super subject" approach to estimate sensitivity (d') and response criterion (c) according to signal detection theory (SDT). Some readers may have issues with this super subject approach, but my main concern is with the lack of precision used by the authors when interpreting the results from this super subject analysis.

Only the super subject had above-chance sensitivity (and it was quite modest, with d' values between 0.07 and 0.51), but the authors over-interpret these results as applying to every subject. The methods and analyses cannot determine if any individual subject could report the features above-chance. Therefore, the following list of quotes should be revised for accuracy or removed from the paper as they are misleading and are not supported by the super subject analysis: "Altogether this approach reveals that subjects can report above-chance the features of stimuli (color, shape, and location) that they had claimed not to notice under traditional yes/no questioning" (p.6)

"In other words, nearly two-thirds of subjects who had just claimed not to have noticed any additional stimulus were then able to correctly report its location." (p.6)

"Even subjects who answer "no" under traditional questioning can still correctly report various features of the stimulus they just reported not having noticed, suggesting that they were at least partially aware of it after all." (p.8)

"Why, if subjects could succeed at our forced-response questions, did they claim not to have noticed anything?" (p.8)

"we found that observers could successfully report a variety of features of unattended stimuli, even when they claimed not to have noticed these stimuli." (p.14)

"our results point to an alternative (and perhaps more straightforward) explanation: that inattentionally blind subjects consciously perceive these stimuli after all... they show sensitivity to IB stimuli because they can see them." (p.16)

"In other words, the inattentionally blind can see after all." (p.17)

We thank the reviewer for pointing out how these quotations may be misleading as regards our central claim. We intended them all to be read generically as concerning the group, and not universally as claiming that all subjects could report above-chance/see the stimuli etc. We agree entirely that the latter universal claim would not be supported by our data. In contrast, we do contend that our super-subject analysis shows that, *as a group*, subjects traditionally

considered intentionally blind exhibit residual sensitivity to features of stimuli (color, shape, and location) that they had all claimed not to notice, and likewise that *as a group* they could succeed at our forced-choice questions.

To ensure this claim is clear throughout the paper, and that we are not interpreted as making an unsupported universal claim we have revised the language in all of the quotations above, as follows, as well as in numerous other places in the paper.

“Altogether this approach reveals that subjects can report above-chance the features of stimuli (color, shape, and location) that they had claimed not to notice under traditional yes/no questioning” (p.6) => “Altogether this approach reveals that as a group subjects can report above-chance the features of stimuli (color, shape, and location) that they had all claimed not to notice under traditional yes/no questioning” (p.6)

“Even subjects who answer “no” under traditional questioning can still correctly report various features of the stimulus they just reported not having noticed, suggesting that they were at least partially aware of it after all.” (p.8) => “... even subjects who answer “no” under traditional questioning can, as a group, still correctly report various features of the stimuli they just reported not having noticed, indicating significant group-level sensitivity to visual features. Moreover, these results are even consistent with an alternative hypothesis about IB, that as a group, subjects who would traditionally be classified as inattentionally blind are in fact at least partially aware of the stimuli they deny noticing.” (p.8)

“Why, if subjects could succeed at our forced-response questions, did they claim not to have noticed anything?” (p.8) => “Why, if subjects could succeed at our forcedresponse questions as a group, did they all individually claim not to have noticed anything?” (p.8)

“we found that observers could successfully report a variety of features of unattended stimuli, even when they claimed not to have noticed these stimuli.” (p.14) => “we found that groups of observers could successfully report a variety of features of unattended stimuli, even when they all individually claimed not to have noticed those stimuli.” (p.14)

“our results point to an alternative (and perhaps more straightforward) explanation: that inattentionally blind subjects consciously perceive these stimuli after all... they show sensitivity to IB stimuli because they can see them.” (p.16) => “our results just as easily raise an alternative (and perhaps more straightforward) explanation: that inattentionally blind subjects may retain a degree of awareness of these stimuli after all.” (p.16) Here deleting: “they show sensitivity to IB stimuli because they can see them.”

“In other words, the inattentionally blind can see after all.” (p.17) => “In other words, as a group, the inattentionally blind enjoy at least some degraded or partial sensitivity to the location, color and shape of stimuli which they report not noticing.” (p.17)

In one case, we felt the sentence was correct as it stood, since it simply reported a fact about our data:

“In other words, nearly two-thirds of subjects who had just claimed not to have noticed any additional stimulus were then able to correctly report its location.” (p.6)

After all, if subjects were entirely blind and simply guessed, it would be true to say that 50% of subjects would be able to correctly report the stimulus location (by guessing).

In addition to these and numerous other changes, we also added the following explicit statement early in the paper to head-off any confusion on this point: “Note that all analyses reported here relate to this super subject as opposed to individual subjects”.

(3) In addition to the d' values for the super subject being slightly above zero, the authors attempted an analysis of response bias to further question the existence of IB. By including in some of their experiments critical trials in which no critical stimulus was presented, but asking subjects the standard Y/N IB question anyway, the authors obtained false alarm and correct rejection rates. When these FA/CR rates are taken into account along with hit/miss rates when critical stimuli were presented, the authors could calculate c (response criterion) for the super subject. Here, the authors report that response criteria are biased towards saying "no, I didn't notice anything". However, the validity of applying SDT to classic Y/N IB questioning is questionable.

For example, with the subject numbers provided in Box 1 (the 2x2 table of hits/misses/FA/CR), one can ask, 'how many subjects would have needed to answer "yes, I noticed something unusual" when nothing was presented on the screen in order to obtain a non-biased criterion estimate, i.e., $c = 0$?' The answer turns out to be 800 subjects (out of the 2761 total subjects in the stimulus-absent condition), or 29% of subjects in this condition.

In the context of these IB paradigms, it is difficult to imagine 29% of subjects claiming to have seen something unusual when nothing was presented. Here, it seems that we may have reached the limits of extending SDT to IB paradigms, which are very different than what SDT was designed for. For example, in classic psychophysical paradigms, the subject is asked to report Y/N as to whether they think a threshold-level stimulus was presented on the screen, i.e., to detect a faint signal in the noise. Subjects complete many trials and know in advance that there will often be stimuli presented and the stimuli will be very difficult to see. In those cases, it seems more reasonable to incorrectly answer "yes" 29% of the time, as you are trying to detect something very subtle that is out there in the world of noise. In IB paradigms, the stimuli are intentionally designed to be highly salient (and unusual), such that with a tiny bit of attention they can be easily seen. When no stimulus is presented and subjects are asked about their own noticing (especially of something unusual), it seems highly unlikely that 29% of them would answer "yes", which is the rate of FAs that would be needed to support the null hypothesis here, i.e., of a non-biased criterion. For these reasons, the analysis of response bias in the current context is questionable and the results claiming to demonstrate a biased criterion do not provide convincing evidence against IB.

We are grateful to the reviewer for highlighting this aspect of our data. We agree with several of these points. For example, it is indeed striking that — given the corresponding hit rate — a false alarm rate of 29% would be needed to obtain an unbiased criterion. At the same time, we would respectfully push back on other points above. In our first experiment that uses the super-subject analysis, for example, d' is 0.51 and highly significant; to describe that figure, as the reviewer does, as "slightly above zero" seemed not quite right to us (and all the more so given that these experiments involve very large samples and preregistered analysis plans).

We also respectfully disagree that our data call into question the validity of applying SDT to classic yes/no IB questioning. The mathematical foundations of SDT are rock solid, and have been applied far more broadly than we have applied them here. In fact, in a way we would suggest that exactly the opposite attitude is appropriate: rather than thinking that IB challenges an immensely well-supported, rigorously tested and broadly applicable mathematical model of perception, we think that the conflict between our SDT-based model of IB and the standard interpretation constitutes strong reason to disfavor the standard interpretation. Several points are worth making here.

First, it is already surprising that 11.03% of our subjects in E2 (46/417) and 7.24% of our subjects in E5 (200/2761) E5 reported noticing a stimulus when no stimulus was present. But

while this may have seemed unlikely in advance of inquiry, this is in fact what the data show and forms the basis of our criterion calculations. Thus, our criterion calculations already factor in a surprising but empirically verified high false alarm rate of subjects answering “yes” when no stimulus was presented and were asked about their noticing. (We also note that the only paper we know of to report a false alarm rate in an IB paradigm, though not one used to calculate a response criterion, found a very consistent false alarm rate of 10.4%. See Devue et al. 2009.)

Second, while the reviewer is of course correct that a common psychophysical paradigm involves detection of a “threshold-level”/faint stimulus in noise, it is widely recognized that SDT has an extremely broad application, being applicable to any situation in which two kinds of event are to be discriminated (Pastore & Scheirer 1975) and being “almost universally accepted as a theoretical account of decision making in research on perceptual detection and recognition and in numerous extensions to applied domains” quite generally (Estes 2002, see also: Wixted 2020). Indeed, cases abound in which SDT has been successfully applied to situations which do not involve near threshold stimuli in noise. To pick two examples at random, SDT has been used in studying acceptability judgments in linguistics (Huang and Ferreira 2020) and the assessment of physical aggression in childstudent interactions (Lerman et al. 2010; for more general discussion of practical applications, see Swets et al. 2000). Given that the framework of SDT is so widely applied and well supported, and that we see no special reason to make an exception, we believe it can be relied on in the present context.

Finally, we note that inattention blindness can in many ways be considered analogous to “near threshold” detection since inattention is precisely thought to degrade or even abolish awareness of stimuli, meaning that our stimuli can be construed as near threshold in the relevant sense. Indeed, our relatively modest d' values suggest that under inattention stimuli are indeed hard to detect. Thus, even were SDT more limited in its application, we think it still would be appropriate to apply to the case of IB.

(4) One of the strongest pieces of evidence presented in the entire paper is the single data point in Figure 3e showing that in Experiment 3, even the super subject group that rated their non-noticing as “highly confident” had a d' score significantly above zero. Asking for confidence ratings is certainly an improvement over simple Y/N questions about noticing, and if this result were to hold, it could provide a key challenge to IB. However, this result hinges on a single data point, it was not replicated in any of the other 4 experiments, and it can be explained by methodological limitations. I strongly encourage the authors (and other readers) to follow up on this result, in an in-person experiment, with improved questioning procedures.

We agree that our finding that even the super-subject group that rated their non-noticing as “highly confident” had a d' score significantly above zero is an especially strong piece of evidence, and we thank the reviewer for highlighting that here. At the same time, we note that while the finding is represented by a single marker in Figure 3e, it seemed not quite right to call this a “single data point”, as the reviewer does, given that it derives from a large pre-registered experiment involving some 7,000 subjects total, with over 200 subjects in the relevant bin — both figures being far larger than a typical IB experiment. It would of course be tremendous to follow up on this result — and we certainly hope our work inspires various follow-up studies. That said, we note that recruiting the necessary numbers of in person subjects would be an absolutely enormous, career-level undertaking — it would involve bringing more than the entire undergraduate population at our own institution, Johns Hopkins, into our laboratory! While those results would obviously be extremely valuable, we wouldn’t want to read the reviewer’s comments as implying that only an experiment of that magnitude — requiring thousands upon thousands of in-person subjects — could make progress on these issues. Indeed, because every subject can only contribute one critical trial

in IB, it has long been recognized as an extremely challenging paradigm to study in a sufficiently well-powered and psychophysically rigorous way. We believe that our large preregistered online approach represents a major leap forward here, even if it involves certain trade-offs.

In the current Experiment 3, the authors asked the standard Y/N IB question, and then asked how confident subjects were in their answer. Asking back-to-back questions, the second one with a scale that pertains to the first one (including a tricky inversion, e.g., "yes, I am confident in my answer of no"), may be asking too much of some subjects, especially subjects paying half-attention in online experiments. This procedure is likely to introduce a sizeable degree of measurement error.

An easy fix in a follow-up study would be to ask subjects to rate their confidence in having noticed something with a single question using an unambiguous scale:

On the last trial, did you notice anything besides the cross?

(1): I am highly confident I didn't notice anything else

(2): I am confident I didn't notice anything else

(3): I am somewhat confident I didn't notice anything else

(4): I am unsure whether I noticed anything else

(5): I am somewhat confident I noticed something else

(6): I am confident I noticed something else

(7): I am highly confident I noticed something else

If we were to re-run this same experiment, in the lab where we can better control the stimuli and the questioning procedure, we would most likely find a d' of zero for subjects who were confident or highly confident (1-2 on the improved scale above) that they didn't notice anything. From there on, the d' values would gradually increase, tracking along with the confidence scale (from 3-7 on the scale). In other words, we would likely find a data pattern similar to that plotted in Figure 3e, but with the first data point on the left moving down to zero d' . In the current online study with the successive (and potentially confusing) retrospective questioning, a handful of subjects could have easily misinterpreted the confidence scale (e.g., inverting the scale) which would lead to a mixture of genuine high-confidence ratings and mistaken ratings, which would result in a super subject d' that falls between zero and the other extreme of the scale (which is exactly what the data in Fig 3e shows).

One way to check on this potential measurement error using the existing dataset would be to conduct additional analyses that incorporate the confidence ratings from the 2AFC location judgment task. For example, were there any subjects who reported being confident or highly confident that they didn't see anything, but then reported being confident or highly confident in judging the location of the thing they didn't see? If so, how many? In other words, how internally (in)consistent were subjects' confidence ratings across the IB and location questions? Such an analysis could help screen-out subjects who made a mistake on the first question and corrected themselves on the second, as well as subjects who weren't reading the questions carefully enough.

As far as I could tell, the confidence rating data from the 2AFC location task were not reported anywhere in the main paper or supplement.

We are grateful to the reviewer for raising this issue and for requesting that we report the confidence rating data from our 2afc location task in Experiment 3. We now report all this data in our Supplementary Materials (see Supplementary Table 3).

We of course agree with the reviewer's concern about measurement error, which is a concern in all experiments. What, then, of the particular concern that some subjects might have misunderstood our confidence question? It is surely impossible in principle to rule out this possibility; however, several factors bear on the plausibility of this interpretation. First, we explicitly labeled our confidence scale (with 0 labeled as 'Not at all confident' and 3 as 'Highly confident') so that subjects would be very unlikely simply to invert the scale. This is especially so as it is very counterintuitive to treat "0" as reflecting *high* confidence. However, we accept that it is a possibility that certain subjects might nonetheless have been confused in some other way.

So, we also took a second approach. We examined the confidence ratings on the 2afc question of subjects who reported being highly confident that they didn't notice anything.

Reassuringly, the large majority of these high confidence "no" subjects (~80%) reported low confidence of 0 or 1 on the 2afc question, and the majority (51%) reported the lowest confidence of 0. Only 18/204 (9%) subjects reported high confidence on both questions.

Still, the numbers of subjects here are small and so may not be reliable. This led us to take a third approach. We reasoned that any measurement error due to inverting or misconstruing the confidence scale should be symmetric. That is: If subjects are liable to invert the confidence scale, they should do so just as often when they answer "yes" as when they answer "no" – after all the very same scale is being used in both cases. This allows us to explore evidence of measurement error in relation to the much larger number of highconfidence "yes" subjects ($N = 2677$), thus providing a much more robust indicator as to whether subjects are generally liable to misconstrue the confidence scale. Looking at the number of such high confidence noticers who subsequently respond to the 2afc question with low-confidence, we found that the number was tiny. Only 28/2677 (1.05%) of highconfidence noticers subsequently gave the lowest level of confidence on the 2afc question, and only 63/2677 (2.35%) subjects gave *either* of the two lower levels of confidence. In this light, we consider any measurement error due to misunderstanding the confidence scale to be extremely minimal.

What should we make of the 18 subjects who were highly confident non-noticers but then only low-confidence on the 2afc question? Importantly, we do not think that these 18 subjects necessarily made a mistake on the first question and so should be excluded. There is no a priori reason why one's confidence criterion in a yes/no question should carry over to a 2afc question. After all, it is perfectly rationally coherent to be very confident that one didn't see anything but also very confident that *if* there was anything to be seen, it was on the left. Moreover, these 18 subjects were *not* all correct on the 2afc question despite their high confidence (4/18 or 22% getting the wrong answer).

Nonetheless, and again reassuringly, we found that the above-chance patterns in our data remained the same even excluding these 18 subjects. We did observe a slight reduction in percent correct and d' but this is absolutely what one should expect since excluding the most confident performers in any task will almost inevitably reduce performance.

In this light, we consider it unlikely that measurement error fully explains the residual sensitivity found even amongst highly confident non-noticers. That said, we appreciate this concern. We now raise the issue and the analysis of high confidence noticers which addresses it in our revised manuscript. We also thank the reviewer for pressing us to think harder about this issue, which led directly to these new analyses that we believed have strengthened the paper.

(5) In most (if not all) IB experiments in the literature, a partial attention and/or full attention trial (or set of trials) is administered after the critical trial. These control trials are very important for validating IB on the critical trial, as they must show that, when attended, the critical stimuli are very easy to see. If a subject cannot detect the critical stimulus on the control trial, one cannot conclude that they were inattentionally blind on the critical trial, e.g., perhaps the stimulus was just too difficult to see (e.g., too weak, too brief, too far in the periphery, too crowded by distractor stimuli, etc.), or perhaps they weren't paying enough attention overall or failed to follow instructions. In the aggregate data, rates of noticing the stimuli should increase substantially from the critical trial to the control trials. If noticing rates are equivalent on the critical and control trials one cannot conclude that attention was manipulated.

It is puzzling why the authors decided not to include any control trials with partial or full attention in their five experiments, especially given their online data collection procedures where stimulus size, intensity, eccentricity, etc. were uncontrolled and variable across subjects. Including such trials could have actually helped them achieve their goal of challenging the IB hypothesis, e.g., excluding subjects who failed to see the stimulus on the control trials might have reduced the inattentional blindness rates further. This design decision should at least be acknowledged and justified (or noted as a limitation) in a revision of this paper.

We acknowledge that other studies in the literature include divided and full attention trials, and that they could have been included in our work as well. However, we deliberately decided not to include such control trials for an important reason. As the referee comments, the main role of such trials in previous work has been to exclude from analysis subjects who failed to report the unexpected stimulus on the divided and/or full attention control trials.

(For example, as Most et al. 2001 write: “Because observers should have seen the object in the full-attention trial (Mack & Rock, 1998), we used this trial as a control ... Accordingly, 3 observers who failed to see the cross on this trial were replaced, and their data were excluded from the analyses.”) As the reviewer points out, excluding such subjects would very likely have ‘helped’ us. However, the practice is controversial. Indeed, in a review of 128 experiments, White et al. 2018 argue that the practice has “problematic consequences” and “may lead researchers to understate the pervasiveness of inattentional blindness”. Since we wanted to offer as simple and demanding a test of residual sensitivity in IB as possible, we thus decided not to use any such exclusions, and for that reason decided not to include divided/full attention trials.

As recommended, we discuss this decision not to include divided/full attention trials and our logic for not doing so in the manuscript. As we explain, not having those conditions makes it more impressive, not less impressive, that we observed the results we in fact did — it makes our results more interpretable, not less interpretable, and so absence of such conditions from our manuscript should not (in our view) be considered any kind of weakness.

(6) In the discussion section, the authors devote a short paragraph to considering an alternative explanation of their non-zero d' results in their super subject analyses: perhaps the critical stimuli were processed unconsciously and left a trace such that when later forced to guess a feature of the stimuli, subjects were able to draw upon this unconscious trace to guide their 2AFC decision. In the subsequent paragraph, the authors relate these results to above-chance forced-choice guessing in blindsight subjects, but reject the analogy based on claims of parsimony.

First, the authors dismiss the comparison of IB and blindsight too quickly. In particular, the results from experiment 3, in which some subjects adamantly (confidently) deny

seeing the critical stimulus but guess a feature at above-chance levels (at least at the super subject level and assuming the online subjects interpreted and used the confidence scale correctly), seem highly analogous to blindsight. Importantly, the analogy is strengthened if the subjects who were confident in not seeing anything also reported not being confident in their forced-choice judgments, but as mentioned above this data was not reported.

Second, the authors fail to mention an even more straightforward explanation of these results, which is that ~8% of subjects misinterpreted the "unusual" part of the standard IB question used in experiments 1-3. After all, colored lines and shapes are pretty "usual" for psychology experiments and were present in the distractor stimuli everyone attended to. It seems quite reasonable that some subjects answered this first question, "no, I didn't see anything unusual", but then when told that there was a critical stimulus and asked to judge one of its features, adjusted their response by reconsidering, "oh, ok, if that's the unusual thing you were asking about, of course I saw that extra line flash on the left of the screen". This seems like a more parsimonious alternative compared to either of the two interpretations considered by the authors: (1) IB does not exist, (2) super-subject d' is driven by unconscious processing. Why not also consider: (3) a small percentage of subjects misinterpreted the Y/N question about noticing something unusual. In experiments 4-5, they dropped the term "unusual" but do not analyze whether this made a difference nor do they report enough of the data (subject numbers for the Y/N question and 2AFC) for readers to determine if this helped reduce the ~8% overestimate of IB rates.

Our primary ambition in the paper was to establish, as our title suggests, residual sensitivity in IB. The ambition is quite neutral as to whether the sensitivity reflects conscious or unconscious processing (i.e. is akin to blindsight as traditionally conceived). We were evidently not clear about this, however, leading to two referees coming away with an impression of our claims that is different than we intended. We have revised our manuscript throughout to address this. But we also want to emphasize here that we take our data primarily to support the more modest claim that there is residual sensitivity (conscious or unconscious) in the group of subjects who are traditionally classified as inattentionally blind. We believe that this claim has solid support in our data.

We do in the discussion section offer one reason for believing that there is residual awareness in the group of subjects who are traditionally classified as inattentionally blind. However, we acknowledge that this is controversial and now emphasize in the manuscript that this claim "is tentative and secondary to our primary finding". We also emphasize that part of our point is dialectical: Inattentional blindness has been used to argue (e.g.) that attention is required for awareness. We think that our data concerning residual sensitivity at least push back on the use of IB to make this claim, even if they do not provide decisive evidence (as we agree) that awareness survives inattention. (Cf. here, Hirshhorn et al. 2024 who take up a common suggestion in the field that awareness is best assessed by using both subjective and objective measures, with claims about lack of awareness ideally being supported by both; our data suggest at a minimum that in IB objective measures do not neatly line up with subjective measures.)

We hope this addresses the referee's concern that we dismiss the "the comparison of IB and blindsight too quickly". We do not intend to dismiss that comparison at all, indeed we raise it because we consider it a serious hypothesis. Our aim is simply to raise one possible consideration against it. But, again, our main claim is quite consistent with sensitivity in IB being akin to "blindsight".

We also agree with the referee that a possible explanation of why some subjects say they do not notice something unusual in IB paradigms, is not because they didn't notice anything but

because they didn't consider the unexpected stimulus sufficiently unusual. However, the reviewer is incorrect that we did not mention this interpretation; to the contrary, it was precisely the kind of concern which led us to be dissatisfied with standard IB methods and so motivated our approach. As we wrote in our main text: "However, yes/no questions of this sort are inherently and notoriously subject to bias... For example, observers might be under-confident whether they saw anything (or whether what they saw counted as unusual); this might lead them to respond "no" out of an excess of caution." On our view, this is exactly the kind of reason (among other reasons) that one cannot rely on yes/no reports of noticing unusual stimuli, even though the field has relied on just these sorts of questions in just this way.

We do not, however, think that this explanation accounts for why all subjects fail to report noticing, nor do we think that it accounts for our finding of above-chance sensitivity amongst non-noticers. This is for two critical reasons. First, whereas the word "unusual" did appear in the yes/no question in our Experiments 1-3, it did not appear in our Experiments 4 and 5 on dynamic IB. (In both cases, we used the exact wording of such questions in the experiments we were basing our work on.) And, of course, we still found significant residual sensitivity amongst non-noticers in Experiments 4 and 5. Second, in relation to our confidence experiment, we think it unlikely that subjects who were highly confident that they did not notice anything unusual only said that because they thought what they had seen was insufficiently unusual. Yet even in this group of subjects who were maximally confident that they did not notice anything unusual, we still found residual sensitivity.

(7) The authors use sub-optimal questioning procedures to challenge the existence of the phenomenon this questioning is intended to demonstrate. A more neutral interpretation of this study is that it is a critique on methods in IB research, not a critique on IB as a manipulation or phenomenon. The authors neglect to mention the dozens of modern IB experiments that have improved upon the simple Y/N IB questioning methods. For example, in Michael Cohen's IB experiments (e.g., Cohen et al., 2011; Cohen et al., 2020; Cohen et al., 2021), he uses a carefully crafted set of probing questions to conservatively ensure that subjects who happened to notice the critical stimuli have every possible opportunity to report seeing them. In other experiments (e.g., Hirschhorn et al., 2024; Pitts et al., 2012), researchers not only ask the Y/N question but then follow this up by presenting examples of the critical stimuli so subjects can see exactly what they are being asked about (recognition-style instead of free recall, which is more sensitive). These follow-up questions include foil stimuli that were never presented (similar to the stimulus-absent trials here), and ask for confidence ratings of all stimuli. Conservative, pre-defined exclusion criteria are employed to improve the accuracy of their IB-rate estimates. In these and other studies, researchers are very cautious about trusting what subjects report seeing, and in all cases, still find substantial IB rates, even to highly salient stimuli. The authors should consider at least mentioning these improved methods, and perhaps consider using some of them in their future experiments.

The concern that we do not sufficiently discuss the range of "improved" methods in IB studies is well-taken. A similar concern is raised by Reviewer #2 (Dr. Cohen). To address the concern, we have added to our manuscript a substantial new discussion of such improved methods. However, although we do agree that these methods can be helpful and may well address some of the methodological concerns which our paper raises, we do not think that they are a panacea. Thus, our discussion of these methods also includes a substantial discussion of the problems and pitfalls with such methods which led us to favor our own simple forced-response and 2afc questions, combined with SDT analysis. We think this approach is superior both to the classic approach in IB studies *and* to the approach raised by the reviewers.

In particular, we have four main concerns about the follow up questions now commonly used in the field:

First, many follow up questions are used not to exclude people from the IB group but to *include* people in the IB group. Thus, Most et al. 2001 asked follow up questions but used these to increase their IB group, only excluding subjects from the IB group if they *both* reported seeing and answered their follow ups incorrectly: “Observers were regarded as having seen the unexpected object if they answered ‘yes’ when asked if they had seen anything on the critical trial that had not been present before and if they were able to describe its color, motion, or shape.” This means that subjects who saw the object but failed to see its color, say, would be treated as inattentionally blind. This has the purpose of inflating IB rates, in exactly the way our paper is intended to critique. So, in our view this isn’t an improvement but rather part of the approach we take issue with.

Second, many follow up questions remain yes/no questions or nearby variants, all of which are subject to response bias. For example, in Cohen’s studies which the reviewer mentions, it is certainly true that “he uses a carefully crafted set of probing questions to conservatively ensure that subjects who happened to notice the critical stimuli have every possible opportunity to report seeing them.” We agree that this improves over a simple yes/no question in some ways. However, such follow up probes nonetheless remain yes/no questions, subject to response bias, e.g.:

- (1) “Did you notice anything strange or different about that last trial?”
- (2) “If I were to tell you that we did something odd on the last trial, would you have a guess as to what we did?”
- (3) “If I were to tell you we did something different in the second half of the last trial, would you have a guess as to what we did?”
- (4) “Did you notice anything different about the colors in the last scene?”

Indeed, follow up questions of this kind can be especially susceptible to bias, since subjects may be reluctant to “take back” their earlier answers and so be conservative in responding positively to avoid inconsistency or acknowledgement of earlier error. This may explain why such follow up questions produce remarkable consistency despite their rather different wording. Thus, Simons and Chabris (1999) report: “Although we asked a series of questions escalating in specificity to determine whether observers had noticed the unexpected event, only one observer who failed to report the event in response to the first question (“did you notice anything unusual?”) reported the event in response to any of the next three questions (which culminated in “did you see a ... walk across the screen?”). Thus, since the responses were nearly always consistent across all four questions, we will present the results in terms of overall rates of noticing.” Thus, while there are undoubtedly merits to these follow ups, they do not resolve problems of bias.

This same basic issue affects the follow up question used in Pitts et al. 2012 which the reviewer mentions. Pitts et al. write: “If a participant reported not seeing any patterns and rated their confidence in seeing the square pattern (once shown the sample) as a 3 or less (1 = least confident, 5 = most confident), she or he was placed in Group 1 and was considered to be inattentionally blind to the square patterns.” The confidence rating follow-up question here remains subject to bias. Moreover, and strikingly, the inclusion criterion used means that subjects who were moderately confident that they saw the square pattern when shown (i.e. answered 3) were counted as inattentionally blind (!). We do not think this is an appropriate inclusion criterion.

The third problem is that follow up questions are often free/open-response. For instance, Most et al. (2005) ask the follow up question: “If you did see something on the last trial that had not been present during the first two trials, what color was it? If you did not see something, please guess.” This is a much more difficult and to that extent less sensitive

question than our binary forced-response/2afc questions. For this reason, we believe our follow up questions are more suitable for ascertaining low levels of sensitivity.

The fourth and final issue is that whereas 2afc questions are criterion free (in that they naturally have an unbiased decision rule), this is in fact not true of n -afc questions in general, nor is it true in general of *delayed n-alternative match to sample* designs. Thus, even when limited response options are given, they are not immune to response biases and so require SDT analysis. Moreover, some such tasks can involve decision spaces which are often poorly understood or difficult to analyze without making substantial assumptions about observer strategy.

This last point (as well as the first) is relevant to Hirshhorn et al. 2024. Hirshhorn et al. write that they “used two awareness measures. Firstly, participants were asked to rate stimulus visibility on the Perceptual Awareness Scale (PAS, a subjective measure of awareness: Ramsøy & Overgaard, 2004), and then they were asked to select the stimulus image from an array of four images (an objective measure: Jakes & Wichmann, 2006).”

While certainly an improvement on simple yes/no questioning, the PAS remains subject to response bias. On the other hand, we applaud Hirshhorn et al.’s use of objective measures in the context of IB which of course our design implements. However, while Hirshhorn et al. 2024 suggest that their task is a spatial 4afc following the recommendation of this design by Jakes & Wichmann (2006), it is strictly a 4-alternative delayed match to sample task, so it is doubtful if it can be considered a preferred psychophysical task for the reasons Jakes & Wichmann offer. Regardless, the more crucial point is that observers in such a task might be biased towards one alternative as opposed to another. Thus, use of d' (as opposed to percent correct as in Hirshhorn et al. 2024) is crucial in assessing performance in such tasks.

For all these reasons, then, while we agree that the field has taken significant steps to move beyond the simple yes/no question traditionally used in IB studies (and we have revised our manuscript to make this clear); we do not think it has resolved the methodological issues which our paper seeks to highlight and address, and we believe that our approach contributes something additional that is not yet present in the literature. We have now revised our manuscript to make these points much more clearly, and we thank the reviewer for prompting these improvements.

Reviewer #2 (Public review):

In this study, Nartker et al. examine how much observers are conscious of using variations of classic inattentional blindness studies. The key idea is that rather than simply asking observers if they noticed a critical object with one yes/no question, the authors also ask follow-up questions to determine if observers are aware of more than the yes/no questions suggest. Specifically, by having observers make forced choice guesses about the critical object, the authors find that many observers who initially said "no" they did not see the object can still "guess" above chance about the critical object's location, color, etc. Thus, the authors claim, that prior claims of inattentional blindness are mistaken and that using such simple methods has led numerous researchers to overestimate how little observers see in the world. To quote the authors themselves, these results imply that "inattentionally blind subjects consciously perceive these stimuli after all... they show sensitivity to IB stimuli because they can see them."

Before getting to a few issues I have with the paper, I do want to make sure to explicitly compliment the researchers for many aspects of their work. Getting massive amounts of data, using signal detection measures, and the novel use of a "super subject" are all important contributions to the literature that I hope are employed more in the future.

We really appreciate this comment and that the reviewer found our work to make these important contributions to the literature. We wrote this paper expecting not everyone to accept our conclusions, but hoping that readers would see the work as making a valuable contribution to the literature promoting an underexplored alternative in a compelling way. Given that this reviewer goes on to express some skepticism about our claims, it is especially encouraging to see this positive feedback up top!

Main point 1: My primary issue with this work is that I believe the authors are misrepresenting the way people often perform inattentional blindness studies. In effect, the authors are saying, "People do the studies 'incorrectly' and report that people see very little. We perform the studies 'correctly' and report that people see much more than previously thought." But the way previous studies are conducted is not accurately described in this paper. The authors describe previous studies as follows on page 3:

"Crucially, however, this interpretation of IB and the many implications that follow from it rest on a measure that psychophysics has long recognized to be problematic: simply asking participants whether they noticed anything unusual. In IB studies, awareness of the unexpected stimulus (the novel shape, the parading gorilla, etc.) is retroactively probed with a yes/no question, standardly, "Did you notice anything unusual on the last trial which wasn't there on previous trials?". Any subject who answers "no" is assumed not to have any awareness of the unexpected stimulus.

If this quote were true, the authors would have a point. Unfortunately, I do not believe it is true. This is simply not how many inattentional blindness studies are run. Some of the most famous studies in the inattentional blindness literature do not simply ask a yes/no question (e.g., the invisible gorilla (Simons et al. 1999), the classic door study where the person changes (Simons and Levin, 1998), the study where observers do not notice a fight happening a few feet from them (Chabris et al., 2011). Instead, these papers consistently ask a series of follow-up questions and even tell the observers what just occurred to confirm that observers did not notice that critical event (e.g., "If I were to tell you we just did XYZ, did you notice that?"). In fact, after a brief search on Google Scholar, I was able to relatively quickly find over a dozen papers that do not just use a yes/no procedure, and instead as a series of multiple questions to determine if someone is inattentionally blind. In no particular order some papers (full disclosure: including my own):

- (1) Most et al. (2005) Psych Review*
- (2) Drew et al. (2013) Psych Science*
- (3) Drew et al. (2016) Journal of Vision*
- (4) Simons et al. (1999) Perception*
- (5) Simons and Levin (1998) Perception*
- (6) Chabris et al. (2011) iPerception*
- (7) Ward & Scholl (2015) Psych Bulletin and Review*
- (8) Most et al. (2001) Psych Science*
- (9) Todd & Marois (2005) Psych Science*
- (10) Fougner & Marois (2007) Psych Bulletin and Review*
- (11) New and German (2015) Evolution and Human Behaviour*

(12) Jackson-Nielsen (2017) *Consciousness and cognition*

(13) Mack et al. (2016) *Consciousness and cognition*

(14) Devue et al. (2009) *Perception*

(15) Memmert (2014) *Cognitive Development*

(16) Moore & Egeth (1997) *JEP:HPP*

(17) Cohen et al. (2020) *Proc Natl Acad Sci*

(18) Cohen et al. (2011) *Psych Science*

This is a critical point. The authors' key idea is that when you ask more than just a simple yes/no question, you find that other studies have overestimated the effects of inattention blindness. But none of the studies listed above only asked simple yes/no questions. Thus, I believe the authors are mis-representing the field. Moreover, many of the studies that do much more than ask a simple yes/no question are cited by the authors themselves! Furthermore, as far as I can tell, the authors believe that if researchers do these extra steps and ask more follow-ups, then the results are valid. But since so many of these prior studies do those extra steps, I am not exactly sure what is being criticized.

To make sure this point is clear, I'd like to use a paper of mine as an example. In this study (Cohen et al., 2020, Proc Natl Acad Sci USA) we used gaze-contingent virtual reality to examine how much color people see in the world. On the critical trial, the part of the scene they fixated on was in color, but the periphery was entirely in black and white. As soon as the trial ended, we asked participants a series of questions to determine what they noticed. The list of questions included:

(1) "Did you notice anything strange or different about that last trial?"

(2) "If I were to tell you that we did something odd on the last trial, would you have a guess as to what we did?"

(3) "If I were to tell you we did something different in the second half of the last trial, would you have a guess as to what we did?"

(4) "Did you notice anything different about the colors in the last scene?"

(5) We then showed observers the previous trial again and drew their attention to the effect and confirmed that they did not notice that previously.

In a situation like this, when the observers are asked so many questions, do the authors believe that "the inattentionally blind can see after all?" I believe they would not say that and the reason they would not say that is because of the follow-up questions after the initial yes/no question. But since so many previous studies use similar follow-up questions, I do not think you can state that the field is broadly overestimating inattention blindness. This is why it seems to me to be a bit of a strawman: most people do not just use the yes/no method.

We appreciate this reviewer raising this issue. As he (Dr. Cohen) states, his “primary issue” concerns our discussion of the broader literature (which he worries understates recent improvements made to the IB methodology), rather than, e.g., the experiments we’ve run. We take this concern very seriously and address it comprehensively here.

A very similar issue is identified by Reviewer #1, comment (7). To review some of what we say in reply to them: To address the concern we have added to our manuscript a substantial new discussion of such improved methods. However, although we do agree that these methods can be helpful and may well address some of the methodological concerns which our paper raises, we do not think that they are a panacea. Thus, our discussion of these methods also includes a substantial discussion of the problems and pitfalls with such methods which led us to favor our own simple forced-response and 2afc questions, combined with SDT analysis. We think this approach is superior both to the classic approach in IB studies *and* to the approach raised by the reviewers.

In particular, we have three main concerns about the follow up questions now commonly used in the field:

First, many follow up questions are used not to *exclude* subjects from the IB group but to *include* subjects in the IB group. Thus, Most et al. (2001) asked follow up questions but used these to increase their IB group, only excluding subjects from the IB group if they *both* reported seeing and failed to answer their follow ups correctly: “Observers were regarded as having seen the unexpected object if they answered ‘yes’ when asked if they had seen anything on the critical trial that had not been present before and if they were able to describe its color, motion, or shape.” This means that subjects who saw the object but failed to describe it in these respects would be treated as inattentionally blind. This is problematic since failure to describe a feature (e.g., color, shape) does not imply a complete lack of information concerning that feature; and even if a subject did lack all information concerning these features of an object, this would not imply a complete failure to see the object. Similarly, Pitts et al. (2012) asked subjects to rate their confidence in their initial yes/no response from 1 = least confident to 5 = most confident, and used these ratings to include in the IB group those who rated their confidence in seeing at 3 or less. This is evidently problematic, since there is a large gap between being under confident that one saw something and being completely blind to it. More generally, using follows up to inflate IB rates in such ways raises precisely the kinds of issues our paper is intended to critique. So in our view this isn’t an improvement but rather part of the approach we take issue with.

Second, many follow up questions remain yes/no questions or nearby variants, all of which are subject to response bias. For example, in the reviewer’s own studies (Cohen et al. 2020, 2011; see also: Simons et al., 1999; Most et al., 2001, 2005; Drew et al., 2013; Memmert, 2014) a series of follow up questions are used to try and ensure that subjects who noticed the critical stimuli are given the maximum opportunity to report doing so, e.g.:

- (1) “Did you notice anything strange or different about that last trial?”
- (2) “If I were to tell you that we did something odd on the last trial, would you have a guess as to what we did?”
- (3) “If I were to tell you we did something different in the second half of the last trial, would you have a guess as to what we did?”
- (4) “Did you notice anything different about the colors in the last scene?”

We certainly agree that such follow up questions improve over a simple yes/no question in some ways. However, such follow up probes nonetheless remain yes/no questions, intrinsically subject to response bias. Indeed, follow up questions of this kind can be especially susceptible to bias, since subjects may be reluctant to “take back” their earlier answers and so be conservative in responding positively to avoid inconsistency or acknowledgement of earlier error. This may explain why such follow up questions produce remarkable consistency despite their rather different wording. Thus, Simons and Chabris (1999) report: “Although we asked a series of questions escalating in specificity to determine

whether observers had noticed the unexpected event, only one observer who failed to report the event in response to the first question (“did you notice anything unusual?”) reported the event in response to any of the next three questions (which culminated in “did you see a ... walk across the screen?”). Thus, since the responses were nearly always consistent across all four questions, we will present the results in terms of overall rates of noticing.” Thus, while there are undoubtedly merits to these follow ups, they do not resolve problems of bias.

It is also important to recognize that whereas 2afc questions are criterion free (in that they naturally have an unbiased decision rule), this is not true of n -afc nor delayed n -alternative match to sample designs in general. Performance in such tasks thus requires SDT analysis – which itself may be problematic if the decision space is not properly understood or requires making substantial assumptions about observer strategy.

Third, and finally, many follow up questions are insufficiently sensitive (especially with small sample sizes). For instance, Todd, Fougne & Marois (2005) used a 12-alternative match-to-sample task (see similarly: Fougne & Marois, 2007; Devue et al., 2009). And Most et al. (2005) asked an open-response follow-up: “If you did see something on the last trial that had not been present during the first two trials, what color was it? If you did not see something, please guess.” These questions are more difficult and to that extent less sensitive than binary forced-response/2afc questions of the sort we use in our own studies – a difference which may be critical in uncovering degraded perceptual sensitivity.

For all these reasons, then, while we agree that the field has taken significant steps to move beyond the simple yes/no question traditionally used in IB studies (and we have revised our manuscript to make this clear); we do not think it has resolved the methodological issues which our paper seeks to highlight and address, and we believe that our approach of using 2afc or forced-response questions combined with signal detection analysis is an important improvement on prior methods and contributes something additional that is not yet present in the literature. We have now revised our manuscript to make these points much clearer.

Other studies that improve on the standard methodology

This reviewer adds something else, however: A very helpful list of 18 papers which include follow ups and that he believes overcome many of the issues we raise in our paper. To just state our reaction bluntly: We are familiar with every one of these papers (indeed, one of them is a paper by one of us!), and while we think these are all very valuable contributions to the literature, it is our view that *none* of these 18 papers resolves the worries that led us to conduct our work.

Here we briefly comment on the relevant pitfalls in each case. We hope this serves to underscore the importance of our methodological approach.

(1) Most et al. (2005) *Psych Review*

Either a 2-item or 5-item questionnaire was used. The 2-item questionnaire ran as follows:

- (1) On the last trial, did you see anything other than the 4 circles and the 4 squares (anything that had not been present on the original two trials)? Yes No
- (2) If you did see something on the last trial that had not been present during the original two trials, please describe it in as much detail as possible.

This clearly does not substantially improve on the traditional simple yes/no question. Moreover, the second question (as well as being open-ended) was used to *include* additional subjects in the IB group, in that participants were counted as having seen the object only if they responded “yes” to Q1 *and in addition* “were able to report at least one accurate detail” in response to Q2. In other words, either a subject says “no” (and is treated as unaware), or

says “yes” and then is asked to prove their awareness, as it were. If anything, this intensifies the concerns we raise, by inflating IB rates.

The 5-item questionnaire looked like this:

(1) On the last trial, did you see anything other than the black and white L's and T's (anything that had not been present on the first two trials)?

(2) If you did see something on the last trial that had not been present during the first two trials, please describe it.

(3) If you did see something on the last trial that had not been present during the first two trials, what color was it? If you did not see something, please guess. (Please indicate whether you did see something or are guessing)

(4) If you did see something during the last trial that had not been present in the first two trials, please draw an arrow on the “screen” below showing the direction in which it was moving. If you did not see something, please guess. (Please indicate whether you did see something or are guessing)

(5) If you did see something during the last trial that had not been present during the first two trials, please circle the shape of the object below [4 shapes are presented to choose from]. If you did not see anything, please guess. (Please indicate whether you did see something or are guessing)

Q5 was not used for analysis purposes. (It suffers from the second issue raised above.) Q1 is the traditional y/n question. Qs 2&3 are open ended. It is unclear how responses to Q4 were analyzed (at the limit it could be considered a helpful, forced-choice question – though it again would suffer from the second issue raised above). However, as noted with respect to the 2-item questionnaire, these responses were not used to exclude people from the IB group but to *include* people in it. So again, this approach does not in any way address the issues we are concerned about, and if anything, only makes them worse.

(2) Drew et al. (2013) *Psych Science*

All follow ups were yes/no: “we asked a series of questions to determine whether they noticed the gorilla: ‘Did the final trial seem any different than any of the other trials?’, ‘Did you notice anything unusual on the final trial?’, and, finally, ‘Did you see a gorilla on the final trial?’”. So, this paper essentially implements the standard methodology we mention (and criticize).

(3) Drew et al. (2016) *Journal of Vision*

Follow up questions were used, but the reported procedure does not provide sufficient details to evaluate them (we are only told: “After the final trial, they were asked: ‘On that last trial of the task, did you notice anything that was not there on previous trials?’ They then answered questions about the features of the unexpected stimulus on a separate screen (color, shape, movement, and direction of movement).”). It is not clear that these follow ups were used to exclude any subjects from the analysis. Finally, given that the unexpected object could be the same color as the targets/distractors, it is clear that biases would have been introduced which would need to be considered (but which were not).

(4) Simons & Chabris (1999) *Perception*

All follow ups were yes/no: “observers were ... asked to provide answers to a surprise series of additional questions. (i) While you were doing the counting, did you notice anything unusual on the video? (ii) Did you notice any- thing other than the six players? (iii) Did you see anyone else (besides the six players) appear on the video? (iv) Did you see a gorilla [woman carrying an umbrella] walk across the screen? After any “yes” response, observers

were asked to provide details of what they noticed. If at any point an observer mentioned the unexpected event, the remaining questions were skipped.” As noted previously, the analyses in fact did not use these questions to exclude subjects since answers were so consistent.

(5) Simons and Levin (1998) *Perception*

This is a change detection paradigm, not a study of inattention blindness. And in any case, one yes/no follow up was used: “Did you notice that I’m not the same person who approached you to ask for directions?”

(6) Chabris et al. (2011) *iPerception*

Two yes/no questions were asked: “we asked whether the subjects had seen anything unusual along the route, and then whether they had seen anyone fighting.” It seems that follow up questions (a request to describe the fight) were asked *only of those who said yes*.

This is in fact a common procedure – follow up questions only being asked of the “yes” group. As discussed, it is sometimes used to *increase* rates of IB, *compounding* the problem we identify in our paper. So this is another example of a follow-up question that makes the problem we identify worse, not better.

(7) Ward & Scholl (2015) *Psych Bulletin and Review*

Two yes/no questions were used: “...observers were asked whether they noticed ‘anything ... that was different from the first three trials’ — and if so, to describe what was different. They were then shown the gray cross and asked if they had noticed it—and if so, to describe where it was and how it moved. Only observers who explicitly reported not noticing the cross were counted as ‘nonnoticers’ to be included in the final sample (N = 100).” In each case, combining the traditional noticing question with a request to describe and identify may have induced conservative response biases in the noticing question, since a subject might consider being able to describe or identify the unexpected stimulus a precondition of giving a positive answer to the noticing question.

(8) Most et al. (2001) *Psych Science*

The same 5-item questionnaire discussed above in relation to Most et al. (2005) was used:

(1) On the last trial, did you see anything other than the black and white L’s and T’s (anything that had not been present on the first two trials)?

(2) If you did see something on the last trial that had not been present during the first two trials, please describe it.

(3) If you did see something on the last trial that had not been present during the first two trials, what color was it? If you did not see something, please guess. (Please indicate whether you did see something or are guessing)

(4) If you did see something during the last trial that had not been present in the first two trials, please draw an arrow on the “screen” below showing the direction in which it was moving. If you did not see something, please guess. (Please indicate whether you did see something or are guessing)

(5) If you did see something during the last trial that had not been present during the first two trials, please circle the shape of the object below [4 shapes are presented to choose from]. If you did not see anything, please guess. (Please indicate whether you did see something or are guessing)

Q5 was not used for analysis purposes. (It suffers from the second issue raised above.) Q1 is the traditional yes/no question. Qs 2&3 are open ended. It is unclear how responses to Q4

were analyzed (at the limit it could be considered a helpful, forced-choice question – though it again would suffer from the second issue raised above). However, as noted with respect to the two item questionnaire in Most et al. 2005, these responses were not used to exclude people from the IB group but to *include* people in it. So again this approach does not in any way address the issues we are concerned about, and if anything only makes them worse.

(9) Todd, Fougne & Marois (2005) *Psych Science*

“participants were probed with three questions to determine whether they had detected the critical stimulus ... The first question assessed whether subjects had seen anything unusual during the trial; they responded “yes” or “no” by pressing the appropriate key on the keyboard. The second question asked participants to select which stimulus they might have seen among 12 possible objects and symbols selected from MacIntosh font databases. The third question asked participants to select the quadrant in which the critical stimulus may have appeared by pressing one of four keys, each of which corresponded to one of the quadrants.”

These follow ups were used to *include* people in the IB group: “In keeping with previous studies (Most et al., 2001), participants were considered to have detected the critical stimulus successfully if they (a) reported seeing an unexpected stimulus *and* (b) correctly selected its quadrant location.” In line with our third point about sensitivity, the object identity test transpired to be “too difficult even under full-attention conditions ... Thus, performance with this question was not analyzed further.”

(10) Fougne & Marois (2007) *Psych Bulletin and Review*

Same exact methods and problems as with Todd & Marois (2005) *Psych Science*, just discussed.

(11) New and German (2015) *Evolution and Human Behaviour*

“After the fourth trial containing the additional experimental stimulus, the participant was asked, “Did you see anything in addition to the cross on that trial?” and which quadrant the additional stimulus appeared in. They were then asked to identify the stimulus in an array which in Experiment 1 included two variants chosen randomly from the spider stimuli and the two needle stimuli. Participants in Experiment 2 picked from all eight stimuli used in that experiment.”

Our second concern about response biases and the need for appropriate SDT analysis of the 4/8 alternative tasks applies to all these questions. We also note that analyses were only performed on groups separately (those who detected/failed to detect, those who located/failed to locate, and those who identified/failed to identify) and on the group which did all three/failed to do any one of the three. Especially in light of the fact that some subjects could clearly detect the stimulus without being able to identify it (e.g.), the most stringent test given our concerns (which were not obviously New and German’s comparative concerns), would be to consider the group which could not detect, identify *or* localize.

(12) Jackson-Nielsen (2017) *Consciousness and cognition*

This is a very interesting example of a follow-up which used a 3-AFC recognition test:

“participants were immediately asked, “which display looks most like what you just saw?” from 3 alternatives”. However, though such an objective test is definitely to be preferred in our view to an open-ended series of probes, the 3-AFC test administered clearly had issues with response biases, as discussed, and actually yielded significantly *below* chance performance in one of the experiments.

(13) Mack et al. (2016) *Consciousness and cognition*

The follow ups here were essentially yes/no combined with an assessment of surprise. Participants were asked to enter letters into a box, and if they did so “were immediately asked by the experimenter whether they had noticed anything different about the array on this last trial and if they did not, they were told that there had been no letters and their responses to that news were recorded. Clearly, if they expressed surprise, this would be compelling evidence that they were unaware of the absence of the letters. Those observers who did not enter letters and realized there were no letters present were considered aware of the absence.” So, this again has all of the same problems we identify, considering subjects unaware because they expressed surprise.

(14) Devue et al. (2009) *Perception*

An 8-alternative task was used. The authors were primarily interested in a comparative analysis and so did not use this task to exclude subjects. We note that an 8 alternative task is very demanding – compare the 12-alternative task used in Todd, Fougner & Marois (2005). There was an attempt to investigate biases in a separate bias trial, however SDT measures were not used.

(15) Memmert (2014) *Cognitive Development*

“After watching the video and stating the number of passes, participants answered four questions (following Simons & Chabris, 1999): (1) While you were counting, did you perceive anything unusual on the video? (2) Did you perceive anything other than the six players? (3) Did you see anyone else (besides the six players) appear on the video? (4) Did you notice a gorilla walk across the screen? After any “yes” reply, children were asked to provide details of what they noticed. If at any point a child mentioned the unexpected event, the remaining questions were omitted.” All of these follow-up questions are yes/no judgments, used to determine awareness in exactly the way we critique as problematic.

(16) Moore & Egeth (1997) *JEP:HPP*

This study (which includes one of us, Egeth, as author) did use forced choice questions. In one case, the question was 2-alternative, in the other it was 4-alternative. In the latter case, SDT would have been appropriate but was not used. In the former case, it may have been that a larger sample would have revealed evidence of sensitivity to the background pattern (as it stood 55% answered the 2-alternative question correctly). Although these results have been replicated, unfortunately the replication in Wood and Simons 2019 used a 6-alternative recognition task and this was not analyzed using SDT. We also note that the task is rather difficult in this study. Wood and Simons report: “Exclusion rates were much higher than anticipated, primarily due to exclusions when subjects failed to correctly report the pattern on the full-attention trial; we excluded 361 subjects, or 58% of our sample.”

(17) Cohen et al. (2020) *Proc Natl Acad Sci*

While this paper improves over a simple yes/no question in some ways, especially in that it used the follow up questions to *exclude* subjects from the unaware (IB) group, the follow up probes nonetheless remain yes/no questions, subject to response bias, e.g.:

- (1) “Did you notice anything strange or different about that last trial?”
- (2) “If I were to tell you that we did something odd on the last trial, would you have a guess as to what we did?”
- (3) “If I were to tell you we did something different in the second half of the last trial, would you have a guess as to what we did?”
- (4) “Did you notice anything different about the colors in the last scene?”

Follow up questions of this kind can be especially susceptible to bias, since subjects may be reluctant to “take back” their earlier answers and so be conservative in responding positively to avoid inconsistency or acknowledgement of earlier error. This may explain why such follow up questions can produce remarkable consistency despite their rather different wording.

(18) Cohen et al. (2011) *Psych Science*

Here are the probes used in this study:

- (1) Did you notice anything different on that trial?
- (2) Did you notice something different about the background stream of images?
- (3) Did you notice that a different type of image was presented in the background that was unique in some particular way?
- (4) Did you see an actual photograph of a natural scene in that stream?
- (5) If I were to tell you that there was a photograph in that stream, can you tell me what it was a photograph of?

Qs 1-4 are yes/no. Q5 is yes/no with an open-ended response. After this, a 5 or 6-alternative recognition test was administered. So again, this faces the same issues, since y/n questions are subject to bias in the way we have described, and many-alternative tests are more problematic than 2afc tests.

In summary

We *really* appreciate the care that went into compiling this list, and we agree that these papers and the improved methods they contain are relevant. But as hopefully made clear above, the approaches in each of these papers simply don't solve the foundational issues our critique is aimed at (though they may address other issues). This is why we felt our new approach was necessary. And we continue to feel this way even after reading and incorporating these comments from Dr. Cohen.

Nevertheless, there is clearly lots for us to do in light of these comments. And so as noted earlier we have now added a very substantial new section to our discussion section to more fairly and completely portray the state of the art in this literature. This is really to our benefit in the end, since we now not only better acknowledge the diverse approaches present, but also set up ourselves to make our novel contribution exceedingly clear.

Main point 2: Let's imagine for a second that every study did just ask a yes/no question and then would stop. So, the criticism the authors are bringing up is valid (even though I believe it is not). I am not entirely sure that above chance performance on a forced choice task proves that the inattentionally blind can see after all. Could it just be a form of subliminal priming? Could there be a significant number of participants who basically would say something like, "No I did not see anything, and I feel like I am just guessing, but if you want me to say whether the thing was to the left or right, I will just 100% guess"? I know the literature on priming from things like change and inattentional blindness is a bit unclear, but this seems like maybe what is going on. In fact, maybe the authors are getting some of the best priming from inattentional blindness because of their large sample size, which previous studies do not use.

I'm curious how the authors would relate their studies to masked priming. In masked priming studies, observers say they did not see the target (like in this study) but still are above chance when forced to guess (like in this study). Do the researchers here think that

that is evidence of "masked stimuli are truly seen" even if a participant openly says they are guessing?

We're grateful to the reviewer for raising this question. As we say in response to Reviewer #1, our primary ambition in the paper is to establish, as our title suggests, residual sensitivity in IB. The ambition is quite neutral as to whether the sensitivity reflects conscious or unconscious processing (i.e. is akin to blindsight as traditionally conceived, or what the reviewer here suggests may be happening in masked priming). Since we were evidently insufficiently clear about this we have revised our manuscript in several places to clarify that we take our data primarily to support the more modest claim that there is residual sensitivity (conscious or unconscious) in the group of subjects who are traditionally classified as inattentionally blind. We believe that this claim has much more solid support in our data than our secondary and tentative suggestion about awareness.

This said, we do consider masked priming studies to be susceptible to the critique that performance may reflect degraded conscious awareness which is unreported because of conservative response criteria. There is good evidence that response criteria tend to be conservative near threshold (Björkman et al. 1993; see also: Railo et al. 2020), including specifically in masked priming studies (Sand 2016, cited in Phillips 2021). So, we consider it a perfectly reasonable hypothesis that subjects who say they feel they are guessing in fact have conscious access to a degraded signal which is insufficient to reach a conservative response criterion but nonetheless sufficient to perform above chance in 2afc detection. Of course, we appreciate that this hypothesis is controversial, so it is not one we argue for in our paper (though we are happy to share our feelings about it here).

Main point 3: My last question is about how the authors interpret a variety of inattentional blindness findings. Previous work has found that observers fail to notice a gorilla in a CT scan (Drew et al., 2013), a fight occurring right in front of them (Chabris et al., 2011), a plane on a runway that pilots crash into (Haines, 1991), and so forth. In a situation like this, do the authors believe that many participants are truly aware of these items but simply failed to answer a yes/no question correctly? For example, imagine the researchers made participants choose if the gorilla was in the left or right lung and some participants who initially said they did not notice the gorilla were still able to correctly say if it was in the left or right lung. Would the authors claim "that participant actually did see the gorilla in the lung"? I ask because it is difficult to understand what it means to be aware of something as salient as a gorilla in a CT scan, but say "no" you didn't notice it when asked a yes/no question. What does it mean to be aware of such important, ecologically relevant stimuli, but not act in response to them and openly say "no" you did not notice them?

Our view is that in such cases, observers may well have a "degraded" percept of the relevant feature (gorilla, plane, fight etc.). But crucially we do not suggest that this percept is sufficient for observers to recognize the object/event as a gorilla, plane, fight etc. Our claim is only that, in our studies at least, observers (as a group) do have enough information about the unexpected stimuli to locate them, and discriminate certain low level features better than chance. Crudely, it may be that subjects see the gorilla simply as a smudge or the plane as a shadowy patch etc. (One of us who is familiar with the gorilla CT scan stimuli notes that the gorilla is in fact rather hard to see even when you know which slide it is on, suggesting that they are not as "salient" as the reviewer suggests!)

More precisely, in the paper we write that in our view perhaps "...unattended stimuli are encoded in a partial or degraded way. Here we see a variety of promising options for future work to investigate. One is that unattended stimuli are only encoded as part of ensemble representations or summary scene statistics (Rosenholtz, 2011; Cohen et al., 2016). Another is that only certain basic "low-level" or "preattentive" features (see Wolfe & Utochkin, 2019 for

discussion) can enter awareness without attention. A final possibility consistent with the present data is that observers can in principle be aware of individual objects and higher-level features under inattention but that the precision of the corresponding representations is severely reduced. Our central aim here is to provide evidence that awareness in inattention blindness is not abolished. Further work is needed to characterize the exact nature of that awareness.” We hope this sheds light on our perspective while still being appropriately cautious not to go too far beyond our data.

Overall: I believe there are many aspects of this set of studies that are innovative and I hope the methods will be used more broadly in the literature. However, I believe the authors misrepresent the field and overstate what can be interpreted from their results. While I am sure there are cases where more nuanced questions might reveal inattention blindness is somewhat overestimated, claims like "the inattentionally blind can see after all" or "Inattentionally blind subjects consciously perceive the stimuli after all" seem to be incorrect (or at least not at all proven by this data).

Once again, we would like to thank this reviewer for his feedback, which obviously comes from a place of tremendous expertise on these issues. We appreciate his assessment that our studies are innovative and that our methodological advances will be of use more broadly. We also hear the reviewer loud and clear about the passages in question, which on reflection we agree are not as central to our case as the other claims we make (regarding residual sensitivity and conservative responding), and so we have now edited them accordingly to refocus our discussion on only those claims that are central and supported. Thank you for making our paper stronger!

Reviewer #3 (Public review):

Summary:

Authors try to challenge the mainstream scientific as well as popularly held view that Inattentional

Blindness (IB) signifies subjects having no conscious awareness of what they report not seeing (after being exposed to unexpected stimuli). They show that even when subjects indicate NOT having seen the unexpected stimulus, they are at above chance level for reporting features such as location, color or movement of these stimuli. Also, they show that 'not seen' responses are in part due to a conservative bias of subjects, i.e. they tend to say no more than yes, regardless of actual visibility. Their conclusion is that IB may not (always) be blindness, but possibly amnesia, uncertainty etc.

We just thought to say that we felt this was a very accurate summary of our claims, and in ways underscore the modesty we had hoped to convey. This is especially true of the reviewer’s final sentence: “Their conclusion is that IB may not (always) be blindness, but possibly amnesia, uncertainty etc.”; as we noted in response to other reviewers, our claim is not that IB doesn’t exist, that subjects are always conscious of the stimulus, etc.; it is only that the cohort of IB subjects show sensitivity to the unattended stimulus in ways that suggest they are not as blind as traditionally conceived. Thank you for reading us as intended!

Strengths:

A huge pool of (25.000) subjects is used. They perform several versions of the IB experiments, both with briefly presented stimuli (as the classic Mack and Rock paradigm), as well as with prolonged stimuli moving over the screen for 5 seconds (a bit like the famous gorilla version), and all these versions show similar results, pointing in

the same direction: above chance detection of unseen features, as well as conservative bias towards saying not seen.

We're delighted that the reviewer appreciated these strengths in our manuscript!

Weaknesses:

Results are all significant but effects are not very strong, typically a bit above chance. Also, it is unclear what to compare these effects to, as there are no control experiments showing what performance would have been in a dual task version where subjects have to also report features etc for stimuli that they know will appear in some trials

The backdrop to the experiments reported here is the “consensus view” (Noah & Mangun, 2020) according to which inattention completely abolishes perception, such that subjects undergoing IB “have no awareness at all of the stimulus object” (Rock et al., 1992) and that “one can have one’s eyes focused on an object or event ... without seeing it at all” (Carruthers, 2015). In this context, we think our findings of significant above-chance sensitivity (e.g., $d' = 0.51$ for location in Experiment 1; chance, of course, would be $d' = 0$ here) are striking and constitute strong evidence against the consensus view. We of course agree that the residual sensitivity is far lower than amongst subjects who noticed the stimulus. For this reason, we certainly believe that inattention has a dramatic impact on perception. To that extent, our data speak in favor of a “middle ground” view on which inattention substantially degrades but crucially does not abolish perception/explicit encoding. We see this as an importantly neglected option in a literature which has overly focused on seen/not seen binaries (see our section ‘Visual awareness as graded’).

Regarding the absence of a control condition, we think those conditions wouldn’t have played the same role in our experiments as they typically play in other experiments. As Reviewer #1 comments, the main role of such trials in previous work has been to *exclude* from analysis subjects who failed to report the unexpected stimulus on the divided and/or full attention control trials. As Reviewer #1 points out, excluding such subjects would very likely have ‘helped’ us. However, the practice is controversial. Indeed, in a review of 128 experiments, White et al. 2018 argue that the practice has “problematic consequences” and “may lead researchers to understate the pervasiveness of inattention blindness”. Since we wanted to offer as simple and demanding a test of residual sensitivity in IB as possible, we thus decided not to use any exclusions, and for that reason decided not to include divided/full attention trials.

As recommended, we discuss this decision not to include divided/full attention trials and our logic for not doing so in the manuscript. As we explain, not having those conditions makes it more impressive, not less impressive, that we observed the results we in fact did — it makes our results more interpretable, not less interpretable, and so absence of such conditions from our manuscript should not (in our view) be considered any kind of weakness.

There are quite some studies showing that during IB, neural processing of visual stimuli continues up to high visual levels, for example, Vandenbroucke et al 2014 doi:10.1162/jocn_a_00530 showed preserved processing of perceptual inference (i.e. seeing a kanizsa illusion) during IB. Scholte et al 2006 doi: 10.1016/j.brainres.2005.10.051 showed preserved scene segmentation signals during IB. Compared to the strength of these neural signatures, the reported effects may be considered not all that surprising, or even weak.

We agree that such evidence of neural processing in IB is relevant to — and perhaps indeed consistent with — our picture, and we’re grateful to the reviewer for pointing out further studies along those lines. Previously, we mentioned a study from Pitts et al., 2012 in which, as

we wrote, “unexpected line patterns have been found to elicit the same Nd1 ERP component in both noticers and inattentionally blind subjects (Pitts et al., 2012).” We have added references to both the studies which the reviewer mentions – as well as an additional relevant study – to our manuscript in this context. Thank you for the helpful addition.

We do however think that our studies are importantly different to this previous work. Our question is whether processing under IB yields representations which are available for explicit report and so would constitute clear evidence of seeing, and perhaps even conscious experience. As we discuss, evidence for this kind of processing remains wanting: “A handful of prior studies have explored the possibility that inattentionally blind subjects may retain some visual sensitivity to features of IB stimuli (e.g., Schnuerch et al., 2016; see also Kreitz et al., 2020, Nobre et al., 2020). However, a recent meta-analysis of this literature (Nobre et al., 2022) argues that such work is problematic along a number of dimensions, including underpowered samples and evidence of publication bias that, when corrected for, eliminates effects revealed by earlier approaches, concluding “that more evidence, particularly from well-powered pre-registered experiments, is needed before solid conclusions can be drawn regarding implicit processing during inattentional blindness” (Nobre et al., 2022).” Our paper is aimed at addressing this question which evidence of neural processing can only speak to indirectly.

Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

(1) Please report all of the data, especially the number of subjects in each experiment that answered Y/N and the numbers of subjects in each of the Y and N groups that guessed a feature correctly/incorrectly on the 2AFC tasks. And also the confidence ratings for the 2AFC task (for comparison with the confidence ratings on the Y/N questions).

We now report all this data in our (revised) Supplementary Materials. We agree that this information will be helpful to readers.

(2) Consider adding a control condition with partial attention (dual task) or full attention (single task) to estimate the rates of seeing the critical stimulus when it's expected.

This is the only recommendation we have chosen not to implement. The reason, as we explain in detail above (especially in response to Reviewer #1 comment 5), is that this would not in fact be a “control condition” *in our studies*, and indeed would only inflate the biases we are concerned with in our work. As the referee comments, the main role of such trials in previous work has been to *exclude* from analysis subjects who failed to report the unexpected stimulus on the divided and/or full attention control trials. And the practice is controversial: Indeed, in a review of 128 experiments, White et al. 2018 argue that the practice has “problematic consequences” and “may lead researchers to *understate* the pervasiveness of inattentional blindness” (emphasis added). So, our choice not to have such conditions ensures an *especially stringent test* of our central claim. Not having those conditions (and their accompanying exclusions) makes our results more interpretable, not less interpretable, and so the absence of such conditions from our manuscript should not (in our view) be considered any kind of weakness.

We have added a paragraph to our “Design and analytical approach” section explaining the logic behind our deliberate decision not to include divided or full attention trials in our experiments. (For even fuller discussion, see our response to Reviewer #1’s comment 5 above.)

(3) Consider revising the interpretations to be more precise about the distinction between the super subject being above chance versus each individual subject who cannot be at chance or above chance because there was only a single trial per subject.

We have now done this throughout the manuscript, as discussed above. We have also added a substantive additional discussion to our “Design and analytical approach” section discussing what should be said about individual subjects in light of our group level data.

This was a very helpful point, and greatly clarifies the claims we wish to make in the paper. Thank you for this comment, which has certainly made our paper stronger.

Reviewer #2 (Recommendations for the authors):

I would be curious to hear the authors' response to two points:

(1) What do they have to say about prior studies that do more than just ask yes/no questions (and ask several follow-ups)? Are those studies "valid"?

A very substantial new discussion of this important point has been added. As you will see above, we comment on every one of the 18 papers this reviewer raised (as well as the general argument made); we contend that while many of these papers improve on past methodology in various ways, most in fact do “just ask yes/no questions”, and none of them makes the methodological advance we offer in our manuscript. However, this discussion has helped us clarify that very advance, and so working through this issue has really helped us improve our paper and make its relation to existing literature that much clearer. Thank you for raising this crucial point.

(2) Do the authors think it is possible that in many cases, people are just guessing about a critical item's location or color and this is at least in part a form of priming?

We have clarified our discussion in numerous places to further emphasize that our main point concerns above-chance sensitivity, not awareness. Given this, we take very seriously the hypothesis that something like priming of a kind sometimes proposed to occur in cases of blindsight or other putative cases of unconscious perception could be what is driving the responses in non-noticers.

Reviewer #3 (Recommendations for the authors):

(1) Control dual task version with expected stimuli would be nice

We have added a paragraph to our “Design and analytical approach” section explaining the logic behind our deliberate decision not to include divided or full attention trials, which would not in fact be a “control” task in our experiments. For full discussion, see our response to Reviewer 3 above, as well as our summary here in the Recommendations for Authors section in responding to Reviewer 1, recommendation (2).

(2) Please do a better job in discussing and introducing experiments about neural signatures during IB.

A discussion of Vandenbroucke et al. 2014 and Scholte et al. 2006 has been added to our discussion of neural signatures in IB, as well as an additional reference to an important early study of semantic processing in IB (Rees et al., 1999). Thank you for these very helpful suggestions!

<https://doi.org/10.7554/eLife.100337.2.sa0>