AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin

ISSN: 0033-2909

2022, Vol. 148, Nos. 5–6, 370–396
https://doi.org/10.1037/bul0000371

A Tale of Two Theories: A Meta-Analysis of the Attention Set and Load Theories of Inattentional Blindness

Brendan T. Hutchinson[1], Kristen Pammer[2], Kavindu Bandara[3], and Bradley N. Jack[1]
[1] Research School of Psychology, Australian National University
[2] School of Psychology, University of Newcastle
[3] School of Psychological Sciences, University of Melbourne

Inattentional blindness (IB), the failure to notice something right in front of you, offers cognitive scientists and practitioners alike a unique means of studying the nature of visual perception. The present meta-analysis sought to provide the first synthesis of the two leading theories of IB—attention set and load theory. We aimed to estimate the magnitude of the effect of each, how they interact, and how task parameters moderate the magnitude of IB summary estimates. We further sought to address several theoretical issues that have persisted within this broad literature. A total of 317 effect sizes from 81 studies that had manipulated attention set or load were synthesized in a multilevel meta-analysis. Results indicated no significant difference between the attention set summary estimate (odds ratio [*OR*] = 3.26, 95% confidence interval [95% CI] [2.33, 4.57]) and the load summary estimate (*OR* = 1.75, 95% CI [1.10, 2.79]). Theoretical moderators included a difference between feature attention sets (*OR* = 5.02, 95% CI [2.95, 8.55]), semantic attention sets (*OR* = 2.64, 95% CI [1.64, 4.25]), and inherent sets (*OR* = 1.90, 95% CI [1.35, 2.68]), while perceptual load (*OR* = 2.55, 95% CI [1.66, 3.92]) and cognitive load (*OR* = 1.67, 95% CI [1.14, 2.44]) were more comparable. The primary task was found as a key task parameter that moderated summary estimates. The attention set summary estimate was moderated by the number of targets and distractors, whereas the load summary estimate was moderated by the full attention (FA) trial exclusion criterion. Analyses indicated any potential publication bias were overall not likely to impact our conclusions. We discuss the implications of results for a conceptual understanding of IB and how the phenomenon can be more reliably studied in future.

Public Significance Statement
Most of us experience inattentional blindness every day, sometimes as trivial errors like failing to notice a grammatical error in a passage of text. Other times, the phenomenon poses more serious concerns, such as missing a pedestrian while driving. The importance of studying inattentional blindness is therefore clear—not only can it tell us something about how the mind and brain work, but it can help inform interventions for the lapses in attention that occur in everyday life. Our analysis examines the two leading theories of inattentional blindness and finds that the most critical component in explaining why the phenomenon occurs is seemingly the relevance of the information to the observer. Our work emphasizes that more research is needed to unpack how these theories interact, as well as how inattentional blindness may be shaped by situational factors such as time and task.

Keywords: inattentional blindness, attention set, load, perception, attention

Supplemental materials: https://doi.org/10.1037/bul0000371.supp

Famously demonstrated by Simons and Chabris (1999), where observers failed to notice an individual in a gorilla suit casually stroll through a scene of a basketball game, inattentional blindness (IB) is the phenomenon in which an observer fails to notice a conspicuous object or event, despite it being within their visual field. Today, IB is studied by psychologists, neuroscientists, and practitioners alike. For the psychologist, IB offers a method for understanding how attention operates. Through careful manipulation of the features of

Brendan T. Hutchinson https://orcid.org/0000-0002-1695-6854
Kristen Pammer https://orcid.org/0000-0002-3183-5268
Kavindu Bandara https://orcid.org/0000-0003-0693-1909
The authors would like to gratefully acknowledge the following authors, in no particular order, for their helpful assistance during the meta-analysis: Dennis Redlich, Carina Kreitz, Katherine Wood, Ricardo Pazzona, Trafton Drew, Rene Marois, Steven Most, Daniel Simons, and Marco Guicciardi. The authors wish to declare no conflicts of interest or funding support for the present article.

Data and code for the meta-analysis can be accessed at https://osf.io/nbvqx/?view_only=3d1f5759a4374158b1eccf09e29b0286.

Correspondence concerning this article should be addressed to Brendan T. Hutchinson, Research School of Psychology, Australian National University, Building 39, University Avenue, Acton, ACT 2601, Australia. Email: brendan.hutchinson@anu.edu.au

the unnoticed object, the context within which it is presented, and the individual differences between observers, psychologists have drawn closer to answering why some objects in our environment tend to capture attention more quickly and forcefully than others. For the neuroscientist, IB offers a means of studying how the brain generates our perception of the visual world. For example, via experimentally inducing IB with simultaneous brain measurement, neuroscientists have edged closer to isolating the neural signature associated with consciousness (for review, see Hutchinson, 2019). For the practitioner, IB offers a method for understanding the safety implications of real-world attentional lapses, such as in driving (Pammer & Blink, 2013; Pammer et al., 2018), aviation (Kennedy et al., 2017), construction (Liao & Chiang, 2016), surgery (Al-Moteri et al., 2018; Greig et al., 2014; Jones & Johnstone, 2017; Paparella, 2013), and police training (Simons & Schlosser, 2017). A better understanding of the rates of IB under such conditions helps inform training and intervention strategies for the potentially dire consequences of the lapses in attention that occur in the real-world, such as missing a pedestrian while driving.

To date, research on IB has been grounded within two different theories: attention set (Folk et al., 1992) and load theory (Lavie, 2005). In the best of circumstances, these views complement one another. In other instances, they lead to differing predictions and conflicting hypotheses. In this article, we aim to bring together empirical knowledge of IB through meta-analysis of research that has studied the phenomenon underpinned by these theories. Previous quantitative syntheses relevant to IB includes work by Kreitz et al. (2020), a recent meta-analysis by Nobre et al. (2020), and earlier meta-analytic work by Gibbs et al. (2016). The former pooled data from their lab's prior studies and reported that correct guessing was above chance in observers who were, in their original analysis, considered inattentionally blind or deaf (Kreitz et al., 2020). Nobre et al. (2020) reached a similar conclusion in their meta-analysis that focused on implicit processing during IB. Gibbs et al. (2016) presented a systematic review of 13 studies examining the related phenomenon change blindness. They reported several factors that moderated the likelihood of change blindness and raised concerns regarding inconsistent methodologies across studies.

While these previous works were highly valuable, they highlight that meta-analytic work on IB more broadly is lacking and that no quantitative synthesis has yet to examine research on the phenomenon incorporating each of its key theoretical constructs. Meta-analysis can provide a less-biased appraisal by bringing these literatures together and quantitatively assessing the weight of evidence for each theory and how they compare. The importance of such an endeavor is underscored by the fact that, for 2 decades, these have remained largely separate bodies of research. Much of the IB literature underpinned by the attention set theory has developed independent of the IB literature underpinned by load theory. As a consequence, advancements and insights made in one literature have not necessarily carried over to the other. This has meant that, at best, research on the phenomenon has not explored interesting theoretical challenges; and at worst, each literature may have been inadvertently biased toward supporting its own theoretical model. Compounding these conceptual issues, differences across studies in task parameters and methodology, which might bias or obfuscate an explanation of the phenomenon, have yet to be systematically explored. A proper examination of these factors is essential and may elucidate why findings range from—as will be seen—theoretically inconsistent to wholly unexpected from a theoretical standpoint. Our aim here is to bring clarity to some of these conceptual and methodological issues, and in turn, facilitate the development of hypotheses and research questions for future by conducting the first meta-analysis investigating each of its key theoretical constructs. Before we proceed, it is pertinent that we first define our key terms.

Defining Attention

As its name suggests, traditionally IB has been thought to occur because an observer's attention to an object (the *critical stimulus*) has been distracted by, or is otherwise directed toward, some other task (the *primary task*; Mack & Rock, 1998). However, a consensus definition of attention has proven notoriously difficult to obtain, an issue that has undoubtedly complicated a clear understanding of IB. Among other definitions, some describe attention based on functional properties of enhancement or inhibition (Wood & Simons, 2017b); others through endogenous or exogenous mechanisms (van Boxtel et al., 2010), or via object-, feature-, or spatial-components (Koivisto et al., 2009). Perhaps the most common definition views attention as a goal-directed mechanism of selection (Nakayama & Martini, 2011). Here, we adopt the following view of attention: to the extent that information within the immediate vicinity of an organism may be given priority, attention is the mechanism by which prioritization occurs. This definition does not stray from the common understanding of attention, yet is flexible enough to allow a number of common delineations to be made (e.g., spatial vs. feature-based), as well as other more tentative distinctions (e.g., enhancement vs. inhibition). Notably, our definition of attention provides adequate operationalization of its key function (prioritization) while remaining suitable to each of the theoretical frameworks that have been employed within the IB literature.

Defining Inattentional Blindness

Because there exist multiple theoretical frameworks of the phenomenon, it is prudent that a definition of IB is sought that remains theoretically agnostic. In our view, a *methodological* definition—one rooted in its use as a method for manipulating awareness—satisfies this requirement while effectively capturing the most iconic experimental occurrences of the phenomenon.

A typical IB experiment proceeds as follows: participants complete a series of computer-based trials of some task (the *primary task*), such as discriminating between the length of the arms of a briefly presented cross. After several trials, an unexpected object (the *critical stimulus*) appears on screen during a trial (the *critical trial*). Participants have no prior knowledge that this will occur. After the critical trial, participants are questioned on whether they noticed anything unusual, usually with a simple yes/no question ("Did you notice anything different on the last trial?"), followed by an open-ended response ("Describe what was different"), and finally a forced-choice question regarding the object's location and/or identity. Participants will usually then return to the primary task in the following trial (the *divided attention trial*), after which they complete one last trial where they are asked *not* to perform the primary task, but instead to simply watch the visual display (the *full attention trial*). After the experiment, participants are classified as a "noticer" or as "inattentionally blind," based on whether sufficient evidence was provided during questioning to infer that they perceived the critical stimulus.

As the critical stimulus is unexpected, participants who did not notice it are typically not able to provide an accurate account of it, precisely because they have no prior knowledge of what the critical stimulus could feasibly have been. Thus, the defining characteristic of IB, and what distinguishes it from other experimental manipulations of consciousness, is that it is elicited when the critical stimulus is not known about in advance. In fact, "surprise" questioning is a requirement for IB to occur, as the primary task alone is typically not sufficient (nor necessary) to induce IB *if* there is prior knowledge that it will be presented (Mack & Rock, 1998).[1] Studies of IB actually use this as an exclusionary criterion: for a participant to be included, they need be able to perceive the critical stimulus once they are informed of its presence (during the *full attention* trial; but see White et al., 2018). Indeed, for this reason, a well-designed IB experiment will also clarify whether participants had any prior knowledge of the phenomenon in the first place, to ensure that those participants are, at the very least, considered for exclusion. It is also for this reason that IB studies will tend to use between-subject designs—because any use of a within-subject design (where a participant contributes more than one data point toward noticing the unexpected critical stimulus) potentially threatens the validity of the "unexpected" nature of the stimulus.

By withholding knowledge of the stimulus's presence, the experimenter renders the stimulus unexpected and its purpose for task execution unclear. It is this manipulation that induces IB when measured through various objective and/or subjective measures of awareness (noting that there are exceptions, such as the driver who looked but failed to see a pedestrian—a point we return to in the discussion). This common element therefore provides an opportune feature by which our definition of IB can hinge. For present purposes then, IB can be defined, absent of theoretical assumptions, as a failure to perceive an unexpected object (the critical stimulus) that is within an observer's visual field.

Theories of Inattentional Blindness

The Attention Set Account

Turning now to the first of IB's theoretical frameworks, "attention set" originates from the contingent capture hypothesis (Folk et al., 1992) and refers to the notion that an observer's top-down attentional control settings govern which components of the task are attended to and which are ignored. Attention set clarifies why cognitive effort is not required when orienting to relevant aspects of a task: attentional control will automatically orient to process stimuli that are "set" as relevant, while other information is ignored. Attention set thus specifies that informational relevancy or "meaning" is critical to whether any given stimulus is subject to IB, as only information considered "relevant" or meaningful will capture attention and hence be perceived. Typically, the relevance of features is not explicitly operationalized by researchers. Rather, relevance is inferred according to task instructions. For example, if participants are to count the number of times black circles bounce against the edge of a visual display, then attention is assumed to be set to at least four features—shape (circles), color (black), number (four), and motion (bouncing across the display). Other hypotheses pertinent to the attention set account include, among others, "animate monitoring" (Calvillo & Jackson, 2014) and "threat superiority" (Gao & Jia, 2017). In these latter cases, relevance is inferred based upon the inherent properties of the visual object. For example, a threatening object, such as a gun or snake, may automatically draw our attention and hence attention is thought be "inherently" set for such objects.

Support for the attention set account comes from the broad finding that similarity is a powerful means of abolishing IB. With few exceptions, noticing of the critical stimulus substantively increases if features of the critical stimulus match those that attention is set to. A classic example comes from Most et al. (2001), who used the paradigmatic dynamic display where participants track multiple moving objects, a task that has since become one of the dominant paradigms in IB research (see Figure 1). Participants were asked to track either four L shapes or T shapes that moved along independent paths and occasionally bounced off the edges of the display. Within several 15-s trials, participants were required to count the number of times the target items bounced within the display and were to ignore the other "distractor" items. During the third trial, an unexpected cross shape—the critical stimulus—entered and traveled from right to left along the horizontal midline of the display for 5 s. Results demonstrated a classic attention set effect for low-level features—in this case luminance—as rates of noticing the cross were proportional to its degree of luminance similarity with the targets.

The Load Theory of Inattentional Blindness

Load theory is the other theoretical framework applied to IB and can be summarized with two propositions: (a) when perceptual processing requirements of a visual task are exceeded, no resources will remain to process, and hence perceive, unexpected or task-irrelevant stimuli and (b) when the cognitive control demands required of a task are sufficiently taxing, prioritization of task-relevant processing will be reduced; therefore, information that is irrelevant but nevertheless competing for attention will be more readily processed. Load theory consequently stresses two means by which a critical stimulus can be subject to IB: perceptual processing capacity limitations—referred to as perceptual load (Lavie, 2005), or optimum cognitive demand—referred to as cognitive load (Chun et al., 2011).

Load theory sits as an intermediary between early and late selection views of attention (Lavie, 2010), suggesting that the degree to which early or late attentional selection occurs is contingent on spare capacity or "resources" (for review, see Lavie et al., 2014). This leads to the position that unexpected and/or irrelevant information will be *automatically* processed via late selection if the resources are available (Lavie, 2010). The load interpretation therefore accounts for the fact that IB still occurs in instances such as when objects overlap the attended region (Rees et al., 1999), as this is predicted to occur if no spare capacity remains. Load theory also makes the assumption that the degree to which other top-down processes modulate IB will be contingent on processing load. In other words, top-down mechanisms associated with task execution, such as those of attentional control settings, will be subject to the

[1] While the primary task is not necessary to induce IB, primary task performance is also commonly assessed to provide some indication as to whether noticing the critical stimulus had an impact on task performance (e.g., which might suggest attention was re-allocated away from the task and toward the critical stimulus) and as an additional manipulation check to ensure the participant was adequately following instructions.

load requirements of a given task. Thus, irrespective of the likelihood of perceiving the critical stimulus according to attention set, load theory implies that conscious visual awareness—and hence IB—will ultimately be subject to capacity limitations.

Literature that has investigated load under conditions of IB suggests that, broadly speaking, visual perception of the critical stimulus is modulated in a manner consistent with what would be expected by load theory. In most studies, greater noticing rates have been observed in the condition where resources are theorized to be leftover (*low* perceptual load) compared with when resources are exhausted (*high* perceptual load; Calvillo & Jackson, 2014; Cartwright-Finch & Lavie, 2007; Koivisto & Revonsuo, 2009; Remington et al., 2014; White & Davies, 2008).

Theoretical Moderators

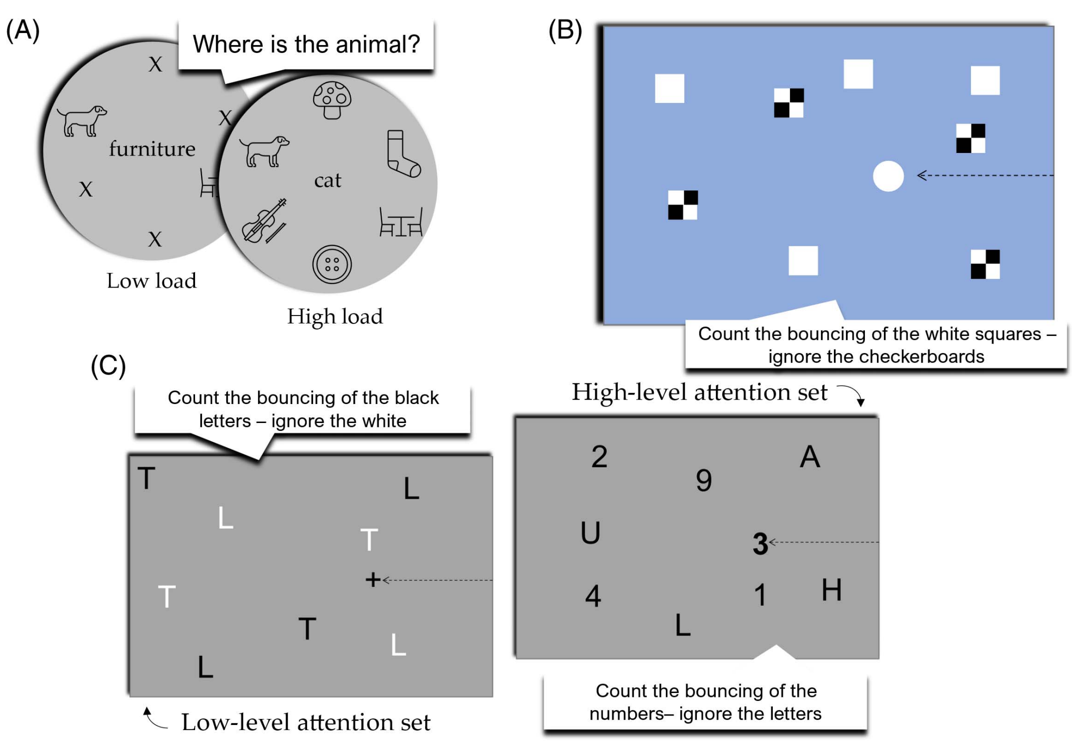
When brought together, each theory appears to offer valid guiding principles for predicting the likelihood of IB. However, close examination shows that many key findings have been left unaccounted for and questions remain unresolved. These issues speak directly to the coherency of each and are consequently central to their validity. In what follows, key challenges of each will be overviewed, with an eye on how meta-analysis can bring clarity to these issues.

How Do Attention Set and Load Interact?

It may not be immediately apparent how the attention set and load accounts of IB conflict. A broad overview of these literatures together suggests that IB—and hence conscious visual awareness of unexpected information—is contingent on both the attentional settings of the observer and the load requirements of the task. There are two issues with this conclusion. The first is that this assumes these theories have been examined conjointly. However, we are aware of less than 10 IB studies to date that have manipulated both theories simultaneously (where visual awareness was the primary outcome measure).

Figure 1

Example Theoretical Moderators of Inattentional Blindness



Note. (A) Attention set and load interaction, where critical stimulus ("cat" in high load) is noticed at equal rates regardless of low or high load. (B) Testing for effect of inhibition. (C) Low-level (feature) and high-level (semantic/categorical) attention set. (A) Adapted with permission from "The Effects of Perceptual Load on Semantic Processing Under Inattention," by M. Koivisto and A. Revonsuo, 2009, *Psychonomic Bulletin & Review*, *16*(5), pp. 864–868 (https://doi.org/10.3758/PBR.16.5.864). Copyright 2009 by Springer Nature. (B) Adapted with permission from "The Role of Similarity in Inattentional Blindness: Selective Enhancement, Selective Suppression, or Both?" by K. Wood and D. J. Simons, 2017, *Visual Cognition*, *25*(9–10), pp. 972–980 (https://doi.org/10.1080/13506285.2017.1365791). Copyright 2017 by Taylor & Francis. (C, left) Adapted with permission from "How Not to Be Seen: The Contribution of Similarity and Selective Ignoring to Sustained Inattentional Blindness," by S. B. Most, D. J. Simons, B. J. Scholl, R. Jimenez, E. Clifford, and C. F. Chabris, *Psychological Science*, *12*(1), pp. 9–17 (https://doi.org/10.1111/1467-9280.00303). Copyright 2001 by SAGE Publications. (C, right) Adapted with permission from "Setting Sights Higher: Category-Level Attentional Set Modulates Sustained Inattentional Blindness," by S. B. Most, 2013, *Psychological Research*, *77*(2), pp. 139–146 (https://doi.org/10.1007/s00426-011-0379-7). Copyright 2009 by Springer Nature. See the online article for the color version of this figure.

The second and more pressing concern is that the sample of studies that have examined both theories conjointly have produced results that cannot be accommodated underneath both theories. The most striking example of this was reported by Koivisto and Revonsuo (2009). These authors simultaneously manipulated both attention set and load, reporting a highly significant attention set effect for semantic relatedness: Words that were semantically relevant to the target items were noticed at a very high rate (91%) compared with those with no semantic relevance (31%). Critically, noticing rates were a mere 7% for semantically unrelated words under high load, while words that were semantically related to targets were perceived at equally high rates (averaging 86.5%) *regardless* of load. Similar findings have been observed for stimuli that are thought to draw attention due to "inherent" relevance (here referred to as an *inherent* set). Beyond threatening or animate objects, characteristic examples of the latter include happy faces (Gupta & Srinivasan, 2015) or the participant's own name (Lin & Yeh, 2014), both of which show comparable rates of noticing under high load and low load conditions.

Collectively, these findings highlight a crucial implication of these distinct theories of IB—each makes opposing predictions under conditions of high load. Load theory would predict that, irrespective of attention set, noticing should decrease under high perceptual load, whereas attention set predicts that noticing should be comparable, irrespective of load, when the critical stimulus matches the observer's attention set. Findings so far appear to support the latter prediction, that noticing can occur *despite* the exhaustion of processing resources when the critical stimulus matches an observer's attention set. This conflicts with the first assumption of load theory, as it reflects perceptual processing of unexpected information despite capacity exhaustion ("late" attentional selection under high perceptual load, Lavie et al., 2014). The reliability of this effect is thus critical to the validity of load theory. Yet, neither its magnitude nor reliability is well understood—a task which meta-analysis is well suited to address.

Similarity to Attended or Ignored Items?

Whilst differences in rates of noticing between attention set "match" and "mismatch" conditions speak to the robustness of the attention set effect, they leave a critical question unanswered: do they arise through similarity to the target items or similarity to the distractor items? Typically, the degree to which the critical stimulus matches with target items is proportional to its similarity with distractor items. Rather than illustrating enhancement of the critical stimulus, these findings may instead reflect inhibition of these same properties due to sharing features that are distracting or irrelevant to the task (Couperus & Lydic, 2019). While attention is commonly viewed as a mechanism of selective enhancement, it also functions via selective inhibition, for example, through the inhibition or suppression of the processing of irrelevant input (Hopf et al., 2006).

In fact, when the degree of similarity between the critical stimulus and target is held constant, and only the relationship between the critical stimulus and distractor items is manipulated, findings suggest that the attention set effect is, at least in part, due to inhibition. An example comes from work by Wood and Simons (2017b). Participants were presented a subset of items that contained color information of both target items and the critical stimulus (checkerboards, see Figure 1). If no inhibition occurred, when the checkerboards were to-be-attended, noticing rates for the critical stimulus should have been equal, irrespective of whether colored black or white, as both were equally similar to the checkerboards. Conversely, when checkerboards were to-be-ignored and participants attended to white squares, there should have been greater noticing when the critical stimulus was white than when it was black. Results were consistent with an account based on both enhancement *and* inhibition: greater noticing for the white critical stimulus when white squares were attended (consistent with enhancement), and lower noticing for the white critical stimulus when checkerboards were attended (consistent with inhibition).

There is consequently reason to suspect that effects observed for the relationship between attention set and IB are, at least in part, a result of selective inhibition (Goldstein & Beck, 2016; Koivisto & Revonsuo, 2008; Most et al., 2001; Wood & Simons, 2017b). While this does not preclude the validity of the attention set account, it complicates the interpretation that IB is an "absence" of attention, precisely because some degree of top-down selection must be instantiated for inhibition to occur (Couperus & Lydic, 2019). Notably, meta-analysis can help clarify the nature of the attention set effect by establishing whether an effect for inhibition can be observed when the overall evidence base is synthesized.

The Types and Levels of Attention Set

One of the more interesting research threads within the attention set literature concerns whether attention is set explicitly or implicitly (here referred to as the *type* of attention set). As previously noted, attention set is usually defined based upon task instructions. Interestingly, literature has suggested the existence of an "implicit" attention set—that is, one not explicitly instructed but either inherently set (e.g., Gupta & Srinivasan, 2015) or inferred to have formed over the course of an experiment. A characteristic example comes from Aimola Davies et al. (2013), who reported noticing rates across several experiments that suggested that participants were more likely to notice objects based on a property *associated* with the target items (e.g., color), rather than the property participants were instructed to pay attention to (e.g., shape). In other words, noticing rates were higher when the critical stimulus was the same color as the target items, even though participants were instructed on which objects to attend according to shape. Perhaps counterintuitively, then, these findings suggest an implicit attention set—one not explicitly instructed—may be stronger than one contingent on explicit task instructions.

This raises an arguably more important question if one considers the fact that the key theoretical prediction of the attention set account—that informational relevancy is crucial to IB—ought to generalize beyond the manipulation of relevance itself. Consider that each theory can, from a practical standpoint, be summarized as a distinctive method for manipulating IB. From this perspective, given some reasonable predefined boundary conditions, we should expect that an attention set effect does *not* require an instruction demarcating the object as (ir)relevant. Otherwise, it is not clear whether the theory offers anything beyond the method used (i.e., an explicit task instruction) to operationalize it. Meta-analysis provides a novel opportunity to critically assess whether this is the case when all IB studies on attention set are pooled.

A similar line of inquiry concerns the different *levels* that attention can be set. Studies often set attention for low-level features such as shape, color, or luminance (Most et al., 2005). But attention is also commonly set for higher level properties of the visual display, such as those contingent on semantic or categorical relationships (Most, 2013; see Figure 1). A logical question that arises from this distinction is whether the attention set effect varies based upon the processing level that attentional tuning operates. An effect with respect to the former is understood to occur because low-level features are processed at relatively early stages of the visual processing hierarchy; hence, attentional enhancement of stimuli that share those features operates from the "bottom up" (Maunsell & Treue, 2006). By comparison, an effect of attentional control for the latter implies a late effect of attention within the visual processing stream, and consequently implicates that unconscious (attentional) processing must operate with some considerable depth. For these reasons, one would expect a feature-based attention set to produce stronger effects than when attention is set at later stage processing levels. Surprisingly, one of the only studies to have made such a comparison found higher rates of noticing with a categorical (i.e., semantic) match compared with a feature match (Koivisto et al., 2004). A quantitative synthesis of the attention set literature provides a direct opportunity to examine—for the first time—whether the attention set effect is moderated by the level that attention is set.

What Is a Load Manipulation, Anyway?

As previously noted, load can be classified according to two "types": perceptual load and cognitive load. Most studies in support of load theory have examined perceptual load, while far fewer have explicitly sought to test predictions made regarding the effects of cognitive load in IB. Of those that have, cognitive load has generally been operationalized as a working memory load manipulation, and opposite effects have been observed (de Fockert & Bremner, 2011; Fougnie & Marois, 2007). Because load theory makes direct (and arguably opposing) predictions regarding these distinguishable types of load, it is essential that a quantitative synthesis seek to examine the summary estimate of each.

It also needs to be pointed out that the distinction between perceptual and cognitive load has, itself, been repeatedly criticized (for review, see Murphy et al., 2016). For example, how does manipulating the speed of targets alter load—is this more reflective of a perceptual or cognitive load manipulation? Beanland and Pammer (2010) manipulated the speed of objects within a dynamic object tracking task. They reported greater noticing in the easy condition (46%) compared with the hard condition (19%). Interestingly, they operationalized the manipulation as one of perceptual load. Yet, tracking multiple objects requires working memory (Allen et al., 2006; Lapierre et al., 2017); thus, this might be better categorized as a cognitive load manipulation. However, in that case, findings would conflict with load theory, if one accepts that higher noticing ought to occur under conditions of high cognitive load (also see Wright et al., 2013).

Questions of this sort are not uncommon and reflect a broader critique that load theory has been subject to concerning the indistinct nature of perceptual and cognitive load. Evidently, it makes little sense to estimate the effect of perceptual or cognitive load without also seeking to understand how important these foundational issues are toward the validity of the distinction between these subconstructs of load. Meta-analysis can provide insight here through assessing how robust the load summary estimate is when it is subject to varying operational definitions of each of the subtypes of load.

Task Parameter and Methodological Moderators

It is possible that what lay at the foundation of theoretical issues is imprecision and inconsistency with respect to how IB itself is studied. IB studies converge on several characteristics: studies almost always employ a primary task to distract participants from noticing the critical stimulus. On the other hand, studies also diverge on many task parameters. In their review of change blindness, Gibbs et al. (2016) suggested that inconclusive findings might be a consequence of different methodologies used. It is therefore imperative that an audit of IB methodologies complements any theoretical investigation of the phenomenon. In what follows, the task parameters we elected to examine are overviewed.

Categorical Moderators

Paradigm

IB is characteristically studied using two different experimental "paradigms"—one where the visual display contains no motion (i.e., is *static*) and one where the visual display has motion (i.e., is *dynamic*; see Figure 2). Though some literature exists examining both paradigms conjointly (e.g., Kreitz, Furley, et al., 2015), none has sought to explicitly subject this task parameter to theoretical cross-examination. A better understanding of how these different paradigms moderate IB is essential for a deeper conceptual understanding of the phenomenon, as it has been speculated to varying degrees that these different paradigms might be "tapping" different types of IB (Most, 2010). Hence, if, for example, attention set studies had predominantly used dynamic paradigms, while load studies had largely used static paradigms, this could present a serious methodological confound to any respective theoretical account of IB.

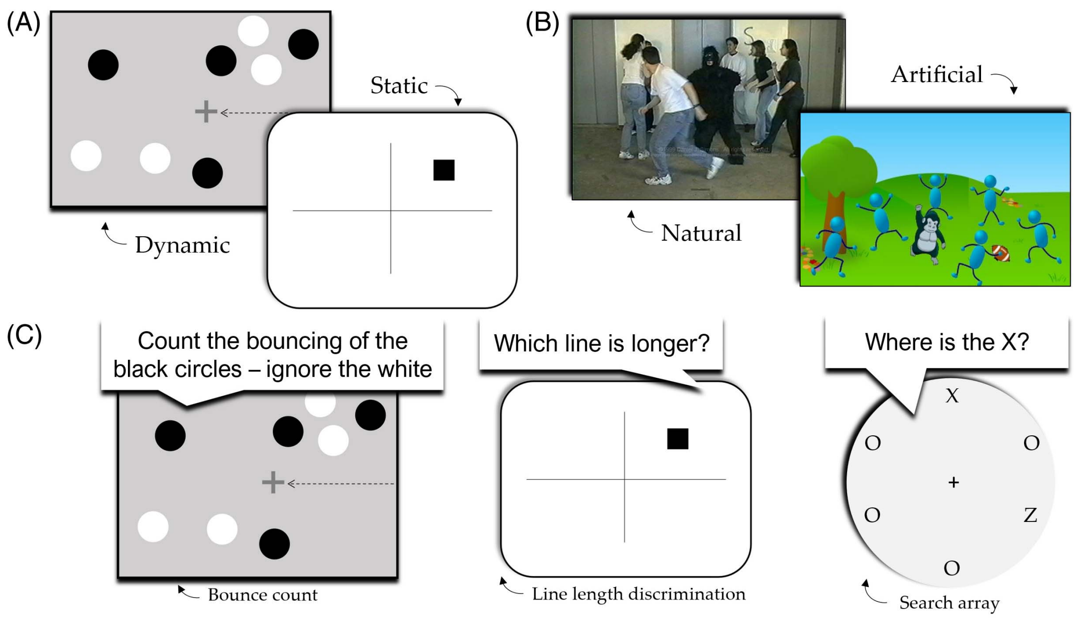
Primary Task

Because visual and executive processes will undoubtedly be taxed to different degrees based on different primary tasks, it is reasonable to posit that theoretical effects might be moderated by the primary task used. Note that paradigm and primary task share some overlap (see Figure 2). For example, bounce count primary tasks are necessarily dynamic (Most et al., 2005). However, primary tasks vary not only within paradigms, but across paradigms too. Therefore, both were considered relevant for examination here.

Setting

IB is appealing to a broad psychological audience, in part, due to its occurrence in real-world settings. Whereas several studies have sought to simulate such settings by implementing "real world" tasks (e.g., Simons & Chabris, 1999), others have opted to study the phenomena using psychophysics-based computer experiments (see Figure 2). Although different results may have emerged through these distinct task settings, for example, because of the stronger control over experimental parameters that the latter would afford over the former, this has remained un-investigated as a potential moderator of theoretical effects. Interestingly, Gibbs et al. (2016)

Figure 2
Example Task Characteristic Moderators of Inattentional Blindness



Note. (A) Static versus dynamic paradigms. (B) Artificial versus natural task setting ("Natural" gorilla provided by Daniel Simons, www.dansimons.com, from Simons & Chabris, 1999; "Artificial" gorilla reproduced with permission from "Toward a Theory of Consciousness: A Review of the Neural Correlates of Inattentional Blindness," by B. T. Hutchinson, 2019, *Neuroscience and Biobehavioral Reviews*, *104*, pp. 87–99, https://doi.org/10.1016/j.neubiorev.2019.06.003). (C) Example of different primary tasks: Bounce count, line length discrimination, and search array. See the online article for the color version of this figure.

reported that studies on change blindness within real-world settings found nonsignificant effects. It is therefore important to examine whether task setting moderates IB summary estimates. A better understanding of this potential task moderator could help speak to the generalizability of these theories to real-world occurrences of the phenomenon.

Quantitative Moderators

The Number of Targets and Distractors

There has been no systematic investigation of whether the number of targets and/or distractors moderates theoretical effects in either literature, despite reason to suspect that these variables may have unique effects on each respective theory. For example, attention set will likely be taxed to a varying degree if observers are required to ignore four moving objects rather than three (but see Koivisto & Revonsuo, 2008). It is important to point out, however, that a common manipulation of load within IB studies involves altering the number of potential targets within a visual display. If rates of noticing differ between conditions of high load and low load in these studies, it stands to reason that load effect sizes would vary with the number of targets. In this case, analysis of whether load effects vary based on the number of objects in the visual display would be redundant. However, perhaps a more crucial benefit of analyzing this task parameter in load studies is that it provides an opportunity to quantify the *difference* between high load and low load across studies. This is especially important given that there is no consistent definition of the difference between these experimental conditions in the load literature. Quantifying this difference within-studies therefore provides an opportunity to clarify the validity of synthesizing load effects across studies.

Critical Stimulus and Trial Duration

It is striking that IB occurs for stimuli that fall within an observer's visual field for several seconds, as a stimulus need only be presented for a matter of milliseconds for it to be detected and identified (Koivisto et al., 2017). In this way, IB demonstrates a conundrum with respect to the intuition that an object should be more likely to be noticed the longer it is presented for. We sought to explicitly test this intuition by assessing how the duration of the critical stimulus and experimental trial impacts upon theoretical effects.

The Number of Precritical Trials

Finally, the number of precritical trials was considered as it might be considered analogous to the effect of practice—with more trials comes more practice on the task; hence, IB would presumably become less likely to occur (Barnhart & Goldinger, 2014; Simons & Jensen, 2009). Surprisingly, this has not been systematically examined in IB (but see Richards et al., 2010); however, the related phenomena *attribute amnesia* and *irrelevance-induced blindness* (for review, see Hutchinson et al., 2022) suggest this impacts upon the likelihood of their occurrence. Meta-analysis therefore gave us the opportunity for a novel examination of whether this task parameter moderates theoretical effects.

Meta-Analytic Objectives

The overall picture to emerge is that theoretical conclusions regarding the nature of IB may be premature without first clarifying the theoretical and methodological questions that have persisted within this broad literature. To mitigate these issues and guide future research, we performed meta-analysis of the attention set and load literatures of IB. Through the pooling of study effect sizes, meta-analysis can estimate a summary effect of a given phenomenon, the external validity of its summary estimate, and its potential moderating variables. Here, we elected to use a multilevel modeling approach where odds ratios (*OR*) were synthesized across attention set and load literatures. Based on the various considerations presented, we had three primary aims: (a) clarify the magnitude of the effect of attention set on IB and the effect of load on IB, and how they statistically compare and interact, (b) test hypotheses concerning potential theoretical moderators of each theory (attention set: inhibition, type, level, load: type), and (c) test hypotheses and explore how task parameter moderators (and study validity check variables, see methods) influence IB summary estimates.

Method

The search strategy employed the Preferred Reporting of Items for Systematic Reviews and Meta-Analyses (Moher et al., 2009). For inclusion, studies must have reported original experimental data. Commentaries, book chapters, and reviews were not included. Where the same data were presented in more than one study, we included the study/publication that more thoroughly described the data. We restricted our analyses to literature on vision (i.e., data on inattentional deafness were not included) and had no inclusion/exclusion criteria based on participant characteristics. We employed three criteria for study inclusion: noticing, prior knowledge, and noticing by theory.

Noticing

Study inclusion required that the study's primary outcome measure was rates of *noticing*, defined as an objective and/or subjective measurement of conscious awareness of the critical stimulus. Objective measures included items such as alternative forced choice questions, whereas subjective measures included open-ended items probing detection or recognition of the critical stimulus. We made no assumptions about what measures were required here as, to date, there is no agreed upon operationalization of IB, and hence, what is required to explicitly distinguish any given participant as a "noticer." Studies were excluded if a measurement of noticing was not included within the study's primary outcome measures. For example, studies that employed reaction time as the dependent variable were not included (e.g., Pugnaghi et al., 2020; Schnuerch et al., 2016; Wright et al., 2013).

Prior Knowledge

An additional requirement for study inclusion was that participants had no prior knowledge and were "surprise" questioned regarding the critical stimulus. This meant that many classic perceptual tasks, for example, visual search or dual-task procedures, were excluded. We must note two issues arose here: (a) where "surprise" questioning impacted prior knowledge and (b) within-subject designs and/or the frequency that the critical stimulus was presented. For example, recognition accuracy for irrelevant face stimuli was employed as the outcome measure in Jenkins et al. (2005) and Srinivasan and Gupta (2010). Both studies employed a dual-tasks procedure where faces were presented many times and were "irrelevant." However, these data were not included because, as pointed out by Jenkins et al. (2005), "[with] repeated immediate testing ... subjects could have known in advance of exposure that faces would become task relevant" (p. 317; but see Supplemental Materials). Repeated questioning is likely sufficient to generate anticipation of the stimuli and "abolish" the IB effect. Thus, while we are aware that some literature has sought to examine the extent to which IB persists *despite* repeated questioning or prior knowledge (e.g., Simons, 2010), where data have included IB rates from repeated questioning (or a within-subject design has otherwise been used, e.g., Beanland et al., 2018; Murphy & Greene, 2016) this has not been included.

Noticing by Theory

Study inclusion required that rates of noticing (as contingent on the first inclusionary criteria) could be compared between conditions as a function of the theoretical construct in question. Studies were included if they directly compared conditions of high versus low load but were excluded if they had no comparison group to assess the effect of load (e.g., Cheng et al., 2019; Wright et al., 2013). For example, Wright et al. (2013) varied perceptual load within-subjects across trials but noticing rates could not be assessed based upon load as the critical stimulus was presented during a high load trial for all participants. This same criterion was employed for attention set studies, where data were included only if a direct comparison could be made between attention set "match" and "mismatch" conditions. A more comprehensive list of inclusion-related issues can be found in the Supplemental Materials.

Literature Search

The primary search for published literature was carried out in two waves. The first occurred in early 2019 across two separate databases using the search terms "inattentional blindness" AND "attention set" OR "load" by the first author. A follow-up search occurred in late 2021 at the request of reviewers. Because the secondary search was more exhaustive, it is the primary one reported here. The secondary search was carried out across six separate databases by the first author. A total of 1,675 records were identified using the search term "inattentional blindness." Following removal of duplicates, the first author screened 981 titles for publication relevance. Here, 894 studies were excluded based on not adhering to the inclusion criteria. The full text was screened for eligibility if the title and/or abstract indicated potential relevance. An additional nine studies were identified for inclusion, either through reference lists of included studies or relevant literature (e.g., reviews), or through our unpublished/gray literature search (see Table 1 and Figure 3).

For unpublished and "gray" literature, following from previous recommendations (Korevaar et al., 2020), we employed three strategies: (a) contact with authors and discipline experts (N = 12, response rate = 50%), (b) search of online conference proceedings and reference guides (N = 4), and (c) search of university theses databases (N = 1). To ensure our search was exhaustive, we also conducted searches for preprint articles on the Open Science Framework and PsyArXiv using the term "inattentional blindness." In total, seven unpublished/gray studies were incorporated into our sample. The final sample included 81 studies: 58 attention set studies and 33 load studies, with 10 studies contributing to counts in both literatures. To gauge interrater reliability for study inclusion, approximately 40% of titles and abstracts, selected at random, were further screened by the third author. Using percent agreement (number of agreement scores/total scores), interrater reliability for study eligibility was 98.6%. Any minor disagreements were resolved through discussion.

Data Collection and Coding

The first author extracted all data and, if required, contacted the corresponding author of an article to request primary outcome missing data. The reliability of coding was gauged in three steps. First, approximately 50% of studies, selected at random, were cross verified by the third author to ensure accuracy in coding. Second, to gauge interrater reliability, an additional 40% of studies distributed approximately equally between both literatures were coded by the third author, blind to the first author's coding. Using percent agreement (number of agreement scores/total scores), interrater reliability was 90.40%. Finally, as an additional reliability check, the entire sample of studies were recoded by the first and third author, distributed approximately equally between authors, prior to the article's publication. Any disagreements in coding were resolved through discussion. The first author (BH) holds a PhD in psychology, while the third author (KB) is a postgraduate student with an Honors degree in psychology.

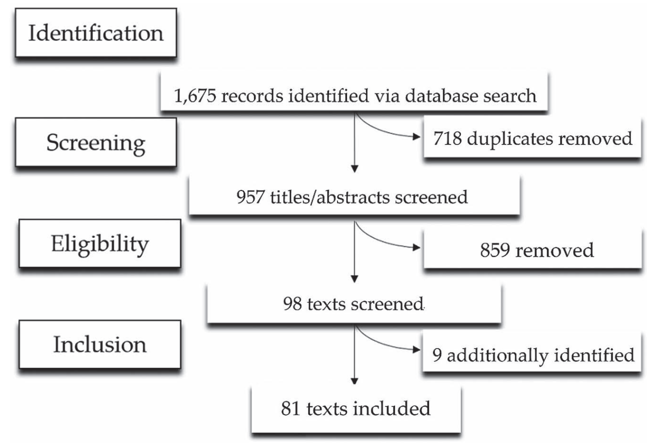
Data extracted included rates of noticing, participant characteristics (mean age, age range, sex), categorical task characteristics (paradigm, setting, task), quantitative task characteristics (number of targets, number of distractors, number of precritical trials, trial duration, critical stimulus duration), theory-related characteristics

Table 1
Literature Search

Search terms	Database	Results: Total	Results: Inclusions
Inattentional blindness	Scopus	485	65
	Pubmed	262	0
	PsycINFO	344	6
	ERIC	15	0
	Philpapers	85	0
	ScienceDirect	484	1
Total		1,675	72

Note. Databases are presented in the order the search was undertaken. Therefore, most studies were derived from Scopus because it was the first database searched.

Figure 3
Flow of Study Reports Into the Research Synthesis



Note. At the screening/eligibility stage, approximately half of all exclusions were due to relevance (e.g., not examining IB), one quarter were due to an insufficient manipulation or outcome measure, one fifth presented no data (i.e., books, commentaries, reviews), and the remainder were excluded for other reasons (e.g., not human participants, could not access). Full-text exclusions (K = 26) included the following: insufficient outcome measure (K = 5), issues with prior knowledge (K = 8), insufficient manipulation (K = 12), and presenting already included data (K = 1). IB = inattentional blindness.

(attention set: set level, set type, inhibition; load: load type), and validity checks (prior knowledge, full attention trial exclusion, poor task performance, load verification, online study—see below; see Table 2).

Quality of studies was assessed using three within-study participant exclusionary criteria: (a) prior knowledge, (b) full attention trial exclusion, and (c) task performance. As previously noted, prior knowledge serves as an exclusionary criterion as it may confound the very nature of IB. Criterion two and three serve as indicators of whether the participant was adequately following instructions or paying attention (but see White et al., 2018). It is almost unanimously the case within published literature that authors are required to justify any inclusions of participants who do *not* satisfy these criteria. For a study to have satisfied these criteria, they need have made explicit mention of whether participants were *considered* for exclusion based upon each. We computed an estimate of study quality based on the sum of how many criteria each study had satisfied ("quality score"), where a score of three reflected that the study had satisfied all three criteria, and a score of zero indicated that the study had satisfied no criteria. In load studies, we also coded for whether authors reported on the verification of the manipulation (i.e., via task performance differences between conditions of high load and low load). Finally, it became apparent during data collection that several studies in the attention set literature were conducted online. This was a methodological consideration, we had not anticipated but chose to code for.

Coding of studies based upon perceptual versus cognitive load (including for sensitivity analysis) is presented in Table 2 of the Supplemental Materials. One issue that arose during the coding of

Table 2
Data Items and Descriptions

Variable	Description
Participant characteristics	Mean age, age range, gender
Theoretical characteristics	
Load	
Perceptual load	As opposed to any selective manipulation of higher level executive processes, manipulation was sensory or perceptual in nature.
Cognitive load	Manipulation was based upon working memory or another executive process.
Attention set level	
Feature	Bottom-up properties such as color, shape, motion, number, location, and action
Semantic	Top-down properties such as object, semantic, and categories
Inherent	Items which are thought to be inherently relevant and thus automatically draw attention (e.g., evolutionary threats)
Inhibition	Whether the role of inhibition was controlled for or otherwise examined.
Attention set type	
Explicit	Attention was "set" explicitly based upon task instructions.
Implicit	Attention "set" was not based upon task instructions but was inferred.
Categorical methodological moderators	
Paradigm	Static (i.e., not moving) versus dynamic (i.e., moving) visual display
Setting	Natural (i.e., real) versus artificial (computer generated) task stimuli/setting
Task	The primary task employed (e.g., bounce count, visual search).
Quantitative methodological moderators	
Target *N*	Number of potential target objects
Distractor *N*	Number of distractor (to-be-ignored) objects
Precritical trials	Number of trials (including practice) preceding the critical trial
Critical stimulus duration	Duration the critical stimulus was presented in the visual display.
Trial duration	Duration of a single trial of the experimental task.
Validity checks	
Prior knowledge	Were participants who had prior knowledge of the task or IB considered for exclusion?
Full attention (FA) trial	Were participants who failed to perceive the critical stimulus in the full attention trial considered for exclusion?
Task performance	Were participants who performed poorly on the primary task considered for exclusion?
Load verification	Whether the load manipulation was verified using, e.g., measures of differences in task performance between groups.
Online study	Was the study conducted online or offline?

Note. Data table of effect sizes, variance estimates, and categorical moderators are presented in Supplemental Table 6. IB = inattentional blindness.

load studies concerned the direction to code cognitive load studies. Discussion of this issue is presented in the Supplemental Materials. In short, all load studies were coded such that effect sizes correspondent with higher rates of noticing in the low load condition would fall on the right side of a forest plot, whereas higher rates of noticing in the high load condition would fall on the left side of a forest plot. Sensitivity analyses were performed where cognitive load studies were reverse-coded, and all summary estimates and analyses were reestimated (refer to Tables 4 and 5 of the Supplemental Materials).

In studies that manipulated load by varying the number of targets or distractors (in particular, static search array studies), we computed a difference score for each based on the normalized target/distractor difference count between conditions of high load and low load. For example, if an experiment used a single potential target in the low load condition and five potential targets in the high load condition, the difference score was four. These "load difference scores" were then used for moderator analysis in the load literature. When performing moderator analysis for "task," we collapsed its coding due to low cluster counts ($k \leq 19$) within some categories. The moderating effect of sex was examined using the proportion of males to females within the sample. Furthermore, for moderator analyses, we computed a critical stimulus exposure time proportion variable based on the proportion of the trial duration that the critical stimulus was presented for (critical stimulus duration/trial duration). Additional notes on coding are presented in Supplemental Materials.

Summary Measures and Results Synthesis

Analyses were conducted using a combination of the *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2022) packages in the R coding environment (R Core Team, 2021). Effect sizes (*OR*s) weighted by inverse sampling variance with 95% confidence intervals (95% CI) were calculated based on rates of noticing (*N* of noticers/total *N* in condition). If rates of noticing were unavailable, alternative statistics to calculate effect sizes were extracted (e.g., chi square). Effect sizes above or below 3.29 *SD*s of the overall mean were considered outlier cases (Tabachnick & Fidell, 2013), while influential cases were identified through their Cook's distance, *DF* betas, and hat values (Viechtbauer & Cheung, 2010).

We submitted the whole sample of studies to a three-level restricted maximum-likelihood random-effects model with robust variance estimation (RVE), where *k* was based upon the number of effects rather than the number of studies (here and throughout, *k* is used to refer to numbers of effects whereas *K* is used to refer to numbers of studies). This approach was advisable in consideration that it can account for dependencies in observations. Specifically, meta-analysis assumes effect sizes are statistically independent. Violation of this assumption can artificially reduce heterogeneity and inflate the risk of Type 1 errors (Cheung, 2014). Pooling of the whole sample of studies resulted in issues of nonindependence. First, many studies were multiexperimental articles and therefore contributed several effect sizes to the overall estimate, a situation that can raise concerns of nonindependence (Van den Noortgate et al., 2013). Second, several studies used multiple treatment and/or control groups, resulting in the possibility of comparing multiple treatment groups to a single control group (or vice versa). Third, the subsample of studies that had examined both theoretical constructs contributed multiple effect sizes that were derived from the same participant sample to the summary estimate. Whereas a "standard" random-effects model incorporates two sources of variance (sampling variability plus between-study heterogeneity), a three-level model can account for dependency through nesting effects within studies to estimate variance at both the within- and between-study level (Van den Noortgate et al., 2013). Estimates of standard error can also be corrected by accounting for correlations between effect sizes from within each study.

We followed recommendations by Pustejovsky and Tipton (2022), according to which our data most closely fit a correlated and hierarchical effects structure. Therefore, we first specified a correlation structure of the effect size estimates within studies using the *impute_covariance_matrix* function from the *clubSandwich* package. Here, effect sizes were clustered, marginal variances were smoothed to the average within each cluster, and an assumed correlation (Spearman's ρ) between effect size estimates within each cluster of 0.6 was used (Pustejovsky & Tipton, 2022). Sensitivity analyses, where all analyses were reperformed using correlation values of 0.4 and 0.8, were also performed (see Supplemental Table 5). Effect sizes were then incorporated into a three-level model where effect sizes were nested within studies, along with the estimated variance covariance matrix, using the *rma.mv* function from *metafor* (Viechtbauer, 2010). This three-level model was then implemented into metafor's *robust* function to apply a cluster-robust adjustment of the model coefficients' variance–covariance matrix with a sandwich-type estimator (see below). This approach was used to model the summary estimate for the whole sample of effect sizes and for all moderator analyses.

Analyses of Moderators

To estimate the unique influence of participant characteristics, theoretical moderators, and methodological characteristics on summary estimates, each moderator variable was submitted as a predictor to separate three-level random effects models. Analysis of moderators one at a time is recommended when the relative importance of different moderators is not known in advance (Assink & Wibbelink, 2016; Cheung, 2014; Viechtbauer, 2010). Furthermore, with three exceptions (see below), we had no prior assumptions regarding interactions between moderators, thus there was little need for complex modeling or interaction terms. The significance of each moderator, including participant characteristics, theoretical moderators, methodological characteristics, study quality checks, and publication bias statistics, was tested using the *robust* function available in *metafor*, which performs an omnibus test based on the *F*-distribution with m and n–p degrees of freedom, where m refers to the number of coefficients, n refers to the number of clusters, and p refers to the number of model coefficients (Viechtbauer, 2010). We applied the "CR2" method as a small-sample correction available via *clubSandwich* in all analyses (Pustejovsky, 2022). Pairwise comparisons were run for each possible comparison for significant moderators with more than two levels using the *anova* function available in *metafor* and applied Bonferroni correction to adjust for familywise error rate with the *p.adjust* function (R Core Team, 2021).

Because theoretical moderators were theory-specific, these were examined on summary estimates in separate three-level meta-analyses for each respective literature. Task parameter moderators were examined at the whole sample level to first establish whether the moderator had an impact on the overall summary estimate. For those variables that were found to be significant, we submitted these to follow-up three-level meta-analysis alongside the variable "literature" (attention set, load). This approach maximized sample size for analysis of each separate moderator and lessened the possibility of Type 1 error by reducing the number of unwieldy cross-analyzing of theoretical and methodological effects (Polanin & Pigott, 2015). At the same time, it allowed us to verify whether methodological characteristics accounted for variance that was otherwise attributed to theoretical effects. This same approach was used to examine the influence of study exclusionary criteria on summary effects, where quality score was submitted as a predictor to a three-level model of the overall summary estimate. However, because there was empirical rationale to understand how different exclusionary criteria might influence summary effects (e.g., White et al., 2018), we also evaluated the unique effect of each validity check in isolation on summary estimates at a literature-specific level.

It is worth noting that we had three prior hypotheses regarding a potential confounding of theoretical and methodological effects. These concerned the relation between (a) paradigm and literature, (b) the number of targets/distractors and attention set, and (c) the difference in the number of targets/distractors (i.e., the "load difference score") between load conditions across studies. We therefore submitted the variables relevant to testing these hypotheses into separate three-level meta-analyses irrespective of the outcomes at the whole sample level.

Heterogeneity

Heterogeneity statistics, including Cochran's Q, I^2, τ^2, and prediction intervals (PI's) for the primary summary estimates, were computed and are presented in Table 5. Cochran's Q provides an indication of whether the observed heterogeneity of the effect size is due to sampling error alone, where a significant Q-value suggests that there is real variance in the effect estimate. I^2 reflects the percentage of the observed variance that can be considered real (not due to sampling error) and was calculated using a manual function written by Mathias Harrer and David Ebert (Harrer et al., 2019). While I^2 can be partitioned based on the proportion of variance attributable at each level, we chose to compute a single I^2 value. Accordingly, I^2 can be interpreted as the proportion of variance attributable to real differences in effect sizes, rather than sampling error (Cheung, 2014). τ^2 or tau-squared refers to the estimated variance of the true effect sizes, and hence can be interpreted as reflecting the variance of the distribution of true outcomes. The PI is based upon ±2 *SD*s of the mean effect size and provides a margin of where the effect will fall in 95% of studies with comparable parameters. We principally relied on and interpreted PI's as a measure of heterogeneity due to their utility—that is, because they can be interpreted to provide an estimation of where the effect size of any future study will lie.

Risk of Bias

A principal concern for any meta-analysis is that the studies included may be a biased sample of all studies that exist on a phenomenon. This might occur for a variety of reasons, such as when studies are not published due to null results (i.e., publication bias, van Aert et al., 2019). To examine such concerns, four methods were employed. First, publication status was examined as a moderator of the overall summary effect using a three-level random effects model with RVE. Second, funnel plots were produced and visually inspected. The classic funnel plot presents a sample of effect sizes against a measure of precision (e.g., standard error, variance, or sample size). If the plot is asymmetric, this can suggest bias because, should none exist, the plot would be expected to be distributed symmetrically and would funnel out around the mean with increasing dispersion corresponding with lower precision of smaller studies (Rothstein et al., 2005).

We produced plots for the overall sample and each literature separately by plotting effect sizes (log *OR*s) against the inverse sample size and quantified the magnitude of asymmetry using a modified variant of Egger's test where the inverse of the sample size is regressed onto the summary estimate. This alternative to Egger's test is recommended as the use of standard error as a measure of precision when examining for publication bias inflates the risk of Type 1 error due to structural dependence between the effect size and their standard error (Rodgers & Pustejovsky, 2021). This can include through biasing regression estimates (Almalik et al., 2021; Deeks et al., 2005; Peters et al., 2006) and inducing trivial asymmetry in funnel plots (Macaskill et al., 2001). The use of (the inverse of) sample size as a measure of precision circumvents this issue and retains Type 1 error rates at the nominal value, irrespective of the magnitude of the summary effect, number of primary studies, or severity of heterogeneity (Macaskill et al., 2001; Moreno et al., 2009; Peters et al., 2006).

Furthermore, it is becoming increasingly recognized that issues of nonindependence need be considered, rather than ignored, when statistically addressing the potential of publication bias within meta-analysis (Rodgers & Pustejovsky, 2021). Therefore, it was appropriate that we assessed for publication bias while accounting for nonindependence in our data. We regressed the inverse of the sample size onto the overall summary effect using a three-level random effects model with RVE, as previously described (Peters et al., 2006; Rodgers & Pustejovsky, 2021). In line with recent recommendations of best practice, the use of a multilevel model allowed us to assess for bias while accounting for statistical nonindependence in effect sizes (Rodgers & Pustejovsky, 2021).

To test the robustness of our results against any potential publication bias that may exist, while accounting for nonindependence, three sensitivity analyses were performed. First, summary estimates and moderator analyses were reestimated based on models "adjusted" for study precision with the simultaneous inclusion of the inverse of the sample size as a covariate. Second, summary estimates and moderator analyses were reestimated using a "trimmed" data set, where the upper quartile of effect sizes based on their inverse sample size were removed (the least precise cases, $k = 83$). Third, we used the *svalue* function from the recently developed *PublicationBias* package to estimate a "worst case" summary estimate (Mathur & VanderWeele, 2020). A worst-case meta-analysis derives a summary estimate only using the "nonaffirmative" effects, and hence provides a hyper conservative pooled estimate under the "worst case scenario" (i.e., assuming all positive effects are due to publication bias). Estimates within *PublicationBias* are calculated via *Robumeta* (Fisher et al., 2017) using the inverse of the sum of the study's variance and allow for clustering of effects (i.e., effect sizes within studies; Mathur & VanderWeele, 2020).

Transparency and Openness

The review protocol was not registered nor published in advance of being conducted. We followed the Preferred Reporting of Items for Systematic Reviews and Meta-Analyses protocol (PRISMA-P) checklist when preparing and the PRISMA reporting guidelines for the final report. The final data set and R code for our analyses are shared and available at https://osf.io/nbvqx/?view_only=3d1f5759a4374158b1eccf09e29b0286.

Results

Sample Characteristics

Eighty-one studies from 16 countries between 1999 and 2021 met the inclusion criteria (including unpublished and gray literature) and contributed 317 effect sizes to the present meta-analysis. One effect size in the overall sample was identified as an outlier and four were identified as influential cases and were removed from all analyses, such that the final sample included 312 effect sizes. The median number of effect sizes per study was 3 (range = 1–21). Data from approximately 19,571 participants were included. Sample sizes ranged from 17 to 295 (*Mdn* = 62). All studies used healthy population samples and only a single study used a sample that might be considered clinically relevant (Swettenham et al., 2014). The most common location of origin was the United States (35.80%), followed by the U.K. (12.35%), and Australia and China (each 9.90%). Table 3 provides descriptive statistics for demographic and categorical moderators. Table 4 provides k and summary statistics broken down by the main theory and task characteristics, as well as associations between categorical and quantitative task parameters. A table of all effect sizes, variance estimates, and categorical task parameters entered into the meta-analysis can be found in Supplemental Table 6. Overall, approximately 46.85% (n = 13,353) of the overall sample of participants were "noticers": 56.65% (n = 6,968) in the treatment group (e.g., attention set match or low load), and 39.41% (n = 6,385) in the control group (e.g., attention set mismatch or high load).[2]

Overall Summary Estimate of the Whole Sample

We were first interested in establishing an overall summary estimate across the whole sample of studies. When all effect sizes were pooled ($k = 312$) using a three-level random-effects model with RVE, the overall summary effect was significant, $OR = 2.68$, 95% CI [2.17, 3.33], $p < .0001$, indicating that substantial differences in rates of IB between experimental groups have been reported within these literatures.

Table 3
Descriptive Statistics for Our Sample of Studies

Variable	Summary statistics
Participants	
Age	M = 25.01, SD = 10.38, range = 4.02–69.60
Gender	43.80% male, SD = 17.08%, range = 0–100%
Location	United States (36%), U.K. (12%), Australia (10%), China (10%), Germany (9%), Finland (5%), Italy, Turkey, Taiwan, the Netherlands, Russia (each 2%), France, India, Ireland, Japan, and New Zealand (each 1%)
Year of publication	*Mdn* = 2014, range = 1999–2021
Publication status	
Published	74 (277)
Unpublished	7 (40)

Note. Location proportions do not add up to 100% due to rounding.

Table 4

Sample Characteristics and Associations Between Categorical and Quantitative Task Parameters

Group	Studies (effect sizes)	CS duration (range)	Trial duration (range)	Precritical trial count (range)	Targets (range)	Distractors (range)
Overall	81 (317)	4.23 s (0.06–60)	33.89 s (0.45–2,570 s)	6.37 (0–176)	2.87 (0–22)	1.88 (0–13)
Theory		$t(93.51) = -1.35$, $p = .18$	$t(244.86) = 1.16$, $p = .25$	$t(71.84) = -2.02$, $p = .05$	$t(103.39) = -1.31$, $p = .19$	$t(135.44) = 1.06$, $p = .29$
Attention set	58 (231)	3.89 s (0.20–15.00 s)	38.95 s (0.45–2,570 s)	4.77 (0–60)	2.75 (0–6)	1.96 (0–4)
Load	33 (86)	5.13 s (0.06–60 s)	20.32 s (0.65–165 s)	11.42 (0–176)	3.20 (0–22)	1.67 (0–13)
Paradigm		$t(205.34) = -15.07$, $p < .0001$	$t(178.03) = -2.69$, $p = .008$	$t(144.18) = 4.07$, $p < .0001$	$t(314.75) = -9.75$, $p < .0001$	$t(188.08) = -28.61$, $p < .0001$
Static	38 (138)	0.78 s (0.06–10 s)	2.97 s (0.45–14.80 s)	10.23 (1–176)	1.77 (0–6)	0.04 (0–2)
Dynamic	45 (179)	6.89 s (0.55–60 s)	57.74 (8.20–2,570 s)	2.84 (0–42.67)	3.73 (0–22)	3.30 (0–13)
Setting		$t(58.33) = 3.42$, $p = .001$	$t(56.01) = 2.12$, $p = .04$	$t(214.61) = -3.14$, $p = .002$	$t(88.79) = -4.22$, $p < .0001$	$t(78.33) = -0.08$, $p = .94$
Natural	19 (57)	7.76 s (0.55–60 s)	143.30 s (2–2,570 s)	3.11 (0–8)	1.91 (0–10)	1.86 (0–13)
Artificial	62 (260)	3.45 s (0.06–10 s)	9.91 s (0.45–45 s)	6.72 (1–176)	3.08 (0–22)	1.88 (0–4)
Task		$\chi^2(3) = 208.43$, $p < .0001$	$\chi^2(3) = 199.99$, $p < .0001$	$\chi^2(3) = 101.88$, $p < .0001$	$\chi^2(3) = 109.48$, $p < .0001$	$\chi^2(3) = 221.15$, $p < .0001$
Count	39 (167)	6.60 s (3–40 s)	56.72 s (8.2–2,570 s)	2.79 (0–42.67)	3.48 (0–4)	3.39 (0–4)
Discrimination	18 (59)	0.32 s (0.1–1 s)	2.04 s (0.65–2.5 s)	9.51 (1–176)	1 (1–1)	0 (0–0)
Search	17 (60)	2.31 s (0.2–60 s)	14.75 s (0.45–600 s)	11.02 (2–125)	2.83 (0–6)	0.43 (0–13)
Other	11 (31)	2.57 s (0.06–10 s)	8.61 s (1.2–22 s)	8 (0–60)	3.26 (0–22)	0.13 (0–2)

Note. Descriptive statistics (mean and range) are presented alongside each subcategory. Inferential statistics are presented at the top row of each variable (theory, paradigm, setting, task) to assess differences in means between groups. Statistics reflect basic *t* tests with a Satterthwaite approximation to the degrees of freedom (for variables with two levels, i.e., theory, paradigm, setting) and Kruskal–Wallis rank sum test (for variables with more than two levels, i.e., task). All statistics are estimated prior to removal of outliers. Counts do not add up for some variables (theory, paradigm, task) due to studies that overlap multiple groups. Durations are presented in seconds. CS = critical stimulus.

Are There Differences in the Mean Effect of Attention Set and Load?

As expected, there was significant heterogeneity in the overall summary estimate, $Q(311) = 2317.93$, $p < .0001$, $I^2 = 81.71$, $\tau^2 = 1.34$, 95% PI [0.26, 27.36]. I^2 suggested that 81.71% of the variance was attributable to real differences in effect sizes, rather than sampling error (Assink & Wibbelink, 2016). The lower bound of the prediction interval indicated that a nonsignificant or null effect cannot be ruled out in future studies. Consequently, there is reason to regard the overall summary estimate as not representative of a single true effect. This is not particularly surprising in consideration that the summary estimate represents a combination of several theoretical constructs.

We proceeded to assess whether there was a significant difference in the mean effect size between literatures. A follow-up three-level model with RVE with literature as a moderator revealed the difference between literatures was close to significant, $F(1, 19.35) = 3.51$, $p = .076$. While the summary estimate for attention set was marginally larger, $OR = 3.26$, 95% CI [2.33, 4.57], $p < .0001$, than the summary estimate for load, $OR = 1.75$, 95% CI [1.10, 2.79], $p = .02$, this did not reach statistical significance. Heterogeneity statistics for the attention set and load summary estimates are presented in Table 5 and indicate that significant heterogeneity was present within both sets of literatures.

How Do Theoretical Moderators Influence the Magnitude of Summary Estimates?

How Do Attention Set and Load Interact?

The first theoretical moderator we sought to examine concerned the nature of the interaction between attention set and load. The sample of studies that had manipulated both theoretical constructs ($K = 9$) were entered into two three-level random-effects models with RVE: Model 1 implemented load (high load, low load) as a moderator on the attention set effect size estimate, while Model 2 implemented attention set (match, mismatch) as a moderator on the load effect size estimate. Neither of these models were significant (both $p > .13$). These models revealed that the attention set summary estimate was significant in both the high load subgroup, $OR = 8.39$, 95% CI [2.37, 29.64], $p = .005$, and low load subgroup, $OR = 3.33$, 95% CI [1.55, 7.12], $p = .007$, whereas the load summary estimate was significant in the mismatch subgroup, $OR = 2.79$, 95% CI [1.03, 7.57], $p = .05$, but not the match subgroup, $OR = 1.51$, 95% CI [0.46, 5.00], $p = .43$.

Does Inhibition Contribute to the Attention Set Effect?

To examine the moderating effect of inhibition on the attention set summary estimate, we first addressed whether there was a difference in effect sizes based on whether the study had ($K = 10$) or had not ($K = 47$) examined inhibition. A three-level model with RVE revealed that this difference was significant, $F(1, 10.98) = 6.74$,

[2] Discrepancies between overall participant N and raw numbers/proportions of IB rates are because the former was calculated after accounting for participants that contributed toward multiple effect sizes. This was not possible for the latter due to IB rates being contingent on the same pool of participants for some effect sizes.

Table 5
Summary Estimates, Heterogeneity Statistics, and Bias-Adjusted Estimates

Model	Effect size estimates			Heterogeneity				
		95% CI					95% PI	
Summary estimate	*OR*	Lower	Upper	*Q*	I^2	τ^2	Lower	Upper
Overall	2.68***	2.17	3.33	2317.93***	81.71	1.34	0.26	27.36
Inverse sample size adjusted	2.16**	1.33	3.49					
Low precision adjusted	2.55***	2.02	3.22					
Worst case adjusted	1.18*	1.01	1.37					
Attention set	3.26***	2.33	4.57	1622.75***	82.91	1.44	0.32	33.31
Inverse sample size adjusted	2.59**	1.44	4.67					
Low precision adjusted	3.19***	2.15	4.74					
Worst case adjusted	1.26*	1.08	1.49					
Load	1.75*	1.10	2.79	437.73***	71.63	0.78	0.16	19.32
Inverse sample size adjusted	1.38	0.79	2.42					
Low precision adjusted	1.58	0.93	2.71					
Worst case adjusted	0.99	0.69	1.43					

Note. Effect size estimates (*OR*s; and their associated 95% CI and 95% PI) are based on estimates from moderator analyses on the overall summary estimate. Heterogeneity statistics (Q, I^2, and τ^2) are based on summary estimates of each literature in separate models. *OR*s reflect the odds of noticing the critical stimulus in the treatment group (e.g., match or low load) compared with the control group (e.g., mismatch or high load). An *OR* of 1 reflects equal rates of noticing between groups. An *OR* greater than 1 indicates an increase in noticing in the treatment group compared with the control group. CI = confidence interval; PI = prediction interval; *OR* = odds ratio.
$^{*}p < .05$. $^{**}p < .005$. $^{***}p < .0005$.

$p = .02$, indicating studies that had examined for the role of inhibition reported larger effect sizes, *OR* = 6.77, 95% CI [2.83, 16.19], $p = .001$, than those that had not, *OR* = 2.46, 95% CI [1.91, 3.17], $p < .0001$.

This was followed by assessing whether an effect for inhibition could be statistically verified. Two comparisons were of interest: the match versus neutral comparison (where the critical stimulus was neither a match with the target category nor distractor category) and the match versus mismatch comparison (where the critical stimulus matched the to-be-ignored distractor category). If IB-related attention set effects were wholly reducible to enhancement when the critical stimulus is a match with the target category, then there should be minimal differences between these comparisons. Any significant difference would indicate a difference in noticing between the neutral and mismatch conditions (relative to the match condition) that would presumably emerge due to the critical stimulus matching the to-be-ignored distractors in the latter condition. The difference in effect size estimates was significant, $F(1, 4.67) = 10.57$, $p = .03$, and in the expected direction, $b = -2.23$, $SE = 0.68$, 95% CI [−4.02, −0.43], indicating a larger difference in noticing between the match and mismatch conditions, *OR* = 21.71, 95% CI [4.53, 104.03], than the match and neutral conditions, *OR* = 2.35, 95% CI [1.05, 5.23].

Does Attention Set Vary by Type or Level?

Two separate three-level models with RVE were performed to evaluate the influence of the moderator's "type" (explicit, implicit) and "level" (feature, semantic, inherent) on the attention set summary estimate. Results indicated that type was significant, $F(1, 34.81) = 15.72$, $p = .0003$, where the overall mean effect size for an explicit attention set was larger, *OR* = 4.62, 95% CI [3.15, 6.78], than that of an implicit attention set, *OR* = 2.00, 95% CI [1.51, 2.64]. Results further indicated that level was significant, $F(2, 29.96) = 5.08$, $p = .01$. Pairwise comparisons indicated the significance of this moderator was driven by a significant difference between a feature-based attention set, *OR* = 5.02, 95% CI [2.95, 8.55], $p < .0001$, and an inherent attention set, *OR* = 1.90, 95% CI [1.35, 2.68], $p = .001$ ($p = .009$ for the comparison, Bonferroni corrected; see Table 6).

Are There Differences in the Mean Effect of Perceptual and Cognitive Load?

A three-level model with RVE with load type (perceptual vs. cognitive) as predictor indicated the difference between summary estimates was not significant, $F(1, 20.85) = 2.62$, $p = .12$. While the summary estimate for perceptual load, *OR* = 2.55, 95% CI [1.66, 3.92], $p = .0003$, was larger than that of cognitive load, *OR* = 1.67, 95% CI [1.14, 2.44], $p = .01$, this did not reach statistical significance (see Table 6). Note that this difference was significant in sensitivity analyses using a different coding scheme ($p = .04$) and with cognitive load reverse-coded ($p = .0002$; refer to Supplemental Materials).

How Do Categorical Task Moderators Influence the Magnitude of the Summary Estimate?

Only a single categorical task moderator was significant when examined in a three-level model of the overall summary estimate with RVE: primary task, $F(3, 25.75) = 3.92$, $p = .02$. We proceeded with a three-level RVE model with the moderator "literature" alongside task, and a model assessing for an interaction between task and literature. Only the former was significant, $F(4, 27.23) = 3.12$, $p = .03$, with the simultaneous effect of "literature" remaining close to significant ($p = .08$). Pairwise comparisons suggested that the significance of this moderator of the overall summary estimate was driven by the difference in effect size estimates for search tasks, *OR* = 3.38, 95% CI [2.51, 4.55], $p < .0001$, compared with discrimination tasks, *OR* = 1.70, 95% CI [1.13, 2.58], $p = .01$ ($p = .03$ for the comparison, Bonferroni corrected).

Table 6
Moderator Analyses Estimates

Model	Point estimate *OR*	95% CI Lower	95% CI Upper
Theoretical moderators			
Attention set*			
Feature based	5.02***	2.95	8.55
Semantic	2.64**	1.64	4.25
Inherent	1.90**	1.35	2.68
Load			
Perceptual	2.55***	1.66	3.92
Cognitive	1.67*	1.14	2.44
Methodological moderators			
Primary task*			
Count	3.14***	2.19	4.51
Search	3.38***	2.51	4.55
Discrimination	1.70*	1.13	2.58
Other	2.15**	1.39	3.34

Note. Effect size estimates (*OR*s; and their associated 95% CI) are based on estimates from moderator analyses on the overall summary estimate. *OR*s reflect the odds of noticing the critical stimulus in the treatment group (e.g., match or low load) compared with the control group (e.g., mismatch or high load). An *OR* of 1 reflects equal rates of noticing between groups. An *OR* greater than 1 indicates an increase in noticing in the treatment group compared with the control group. CI = confidence interval; *OR* = odds ratio.
* $p < .05$. ** $p < .005$. *** $p < .0005$.

How Do Quantitative Task Moderators Influence the Magnitude of the Summary Estimate?

The number of targets was a significant moderator of the attention set summary estimate, $F(1, 24.83) = 8.77$, $b = 0.17$, $SE = 0.06$, $p = .007$, 95% CI [0.05, 0.28], as was the number of distractors, $F(1, 28.26) = 6.76$, $b = 0.20$, $SE = 0.08$, $p = .01$, 95% CI [0.04, 0.36]. Figure 4 shows that there was a tendency for attention set effect sizes to be larger with an increasing number of potential targets or distractors.

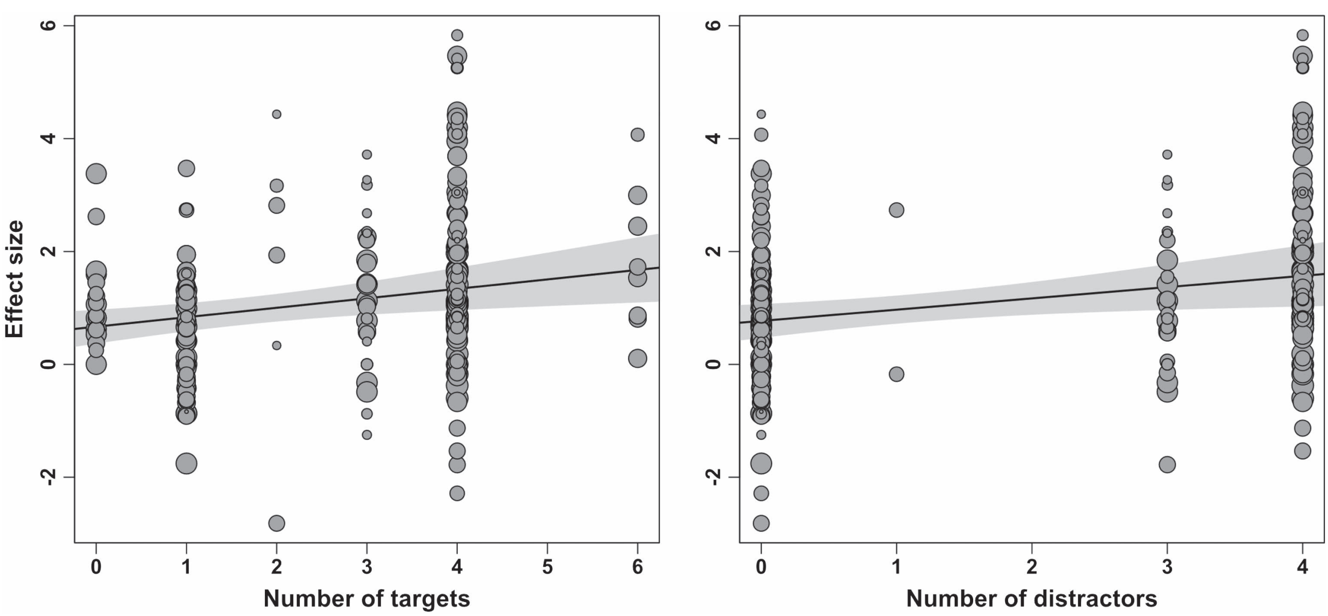
When examined in a three-level model with RVE, the overall summary estimate was also significantly moderated by the critical stimulus to trial duration proportion variable, $F(1, 26.41) = 6.61$, $b = 1.54$, $SE = 0.60$, $p = .02$, 95% CI [0.31, 2.78]. We proceeded to submit this variable into a three-level model with RVE alongside "literature." This model was significant, $F(2, 25.73) = 3.98$, $p = .03$, indicating that, once the effect of literature was considered ($p = .10$), the critical stimulus to trial duration proportion remained marginally significant, $b = 1.22$, $SE = 0.61$, $p = .055$, 95% CI [−0.03, 2.46].

To explore this further, we examined the moderating effect of critical stimulus to trial duration proportion in separate three-level models with RVE for each literature (note that a three-level RVE model specifying an interaction term between critical stimulus proportion and literature was significant, $p = .05$). This indicated that the critical stimulus to trial duration proportion was significant only in the attention set literature, $F(1, 18.86) = 9.31$, $b = 2.08$, $SE = 0.68$, $p = .007$, 95% CI [0.65, 3.51]. Figure 5 shows that there was a tendency for attention set effect sizes to be larger with a longer critical stimulus to trial duration proportion.

How Does Study Quality Influence the Magnitude of the Summary Estimate?

When examined in a three-level model with RVE, the effect of quality score was not a significant moderator of the overall summary estimate ($p = .11$); however, exploratory analyses suggested that there was a significant difference between studies that scored a three out of three, $OR = 4.74$, 95% CI [2.02, 11.11], $p = .005$, compared with

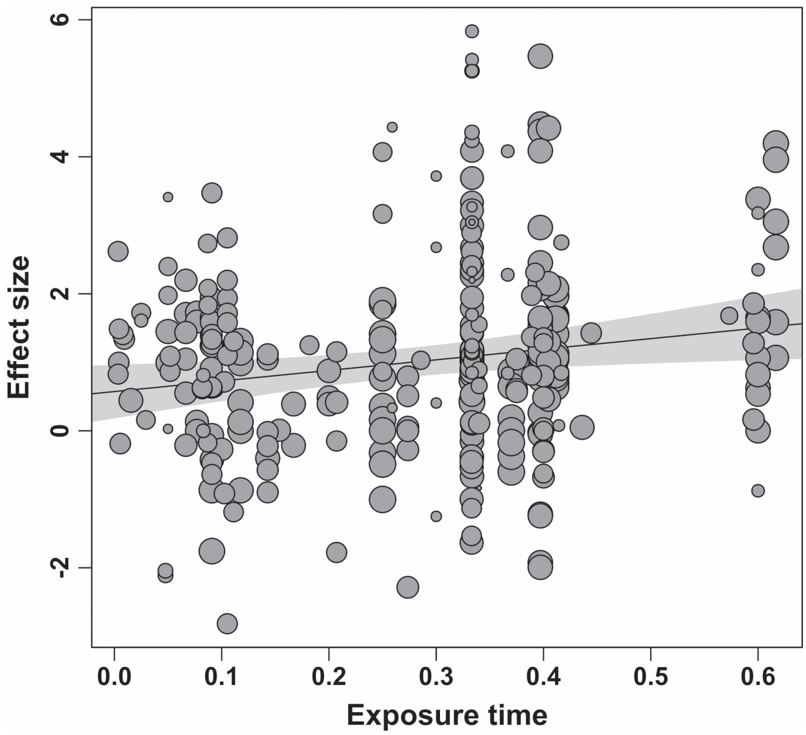
Figure 4
Attention Set Effect Sizes (Log Odds Ratio) as a Function of Number of Targets (Left) and Number of Distractors (Right)



Note. Effect size (y-axis) reflects log odds ratios (*OR*), which can be interpreted as the odds of noticing the critical stimulus in the treatment group (e.g., attention set match) compared with the control group (e.g., attention set mismatch). A log *OR* of 0 reflects equal rates of noticing between groups. A log *OR* greater than 0 indicates an increase in noticing in the treatment group compared with the control group. The size of each point is proportional to its weight. Shaded region reflects confidence interval bounds.

Figure 5

Attention Set Effect Sizes (Log Odds Ratio) as a Function of the Critical Stimulus Exposure Duration (as a Proportion of Trial Duration)



Note. Effect size (y-axis) reflects log odds ratios (*OR*), which can be interpreted as the odds of noticing the critical stimulus in the treatment group (e.g., attention set match) compared with the control group (e.g., attention set mismatch). A log *OR* of 0 reflects equal rates of noticing between groups. A log *OR* greater than 0 indicates an increase in noticing in the treatment group compared with the control group. The size of each point is proportional to its weight. Shaded region reflects confidence interval bounds.

studies that scored a zero out of three, $OR = 1.97$, 95% CI [1.50, 2.60], $p < .0001$ ($p = .04$ for the comparison). The direction of the effect was in the expected direction, such that effect sizes were smaller in studies that scored a zero out of three, $b = -0.65$, $SE = 0.26$, $p = .04$, compared with the summary estimate for studies that scored a three out of three.

The only study validity check variable that was significant in three-level RVE analyses was the full attention trial criterion in the load literature, $F(1, 18.71) = 7.64$, $b = 0.62$, $SE = 0.23$, $p = .01$, 95% CI [0.15, 1.09]. There was a tendency for load studies that satisfied the full attention trial validity check to have larger effect sizes, $OR = 3.25$, 95% CI [1.98, 5.36], $p < .0001$, than those that did not, $OR = 1.75$, 95% CI [1.26, 2.43], $p = .003$.

Are Summary Estimates Impacted by Publication Bias?

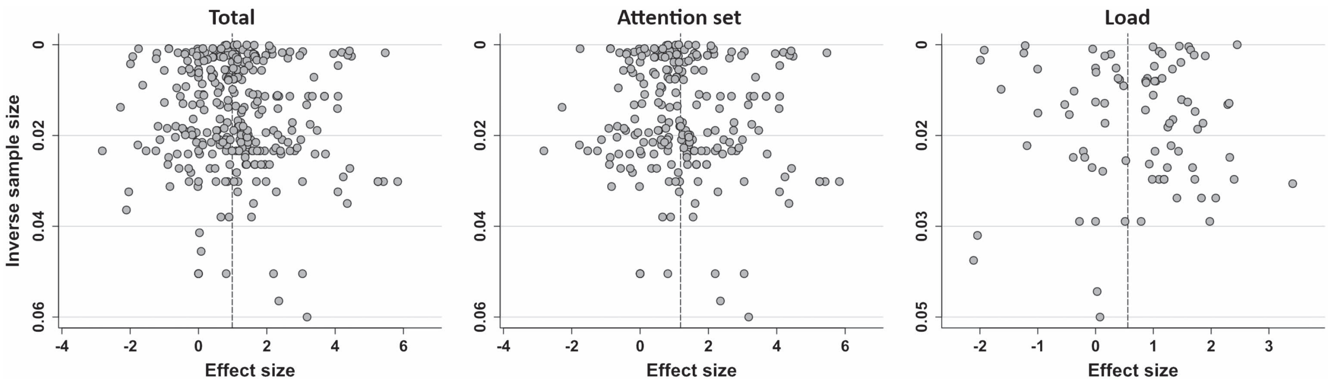
A three-level model with RVE indicated the overall summary estimate was not moderated by publication status, $F(1, 4.86) = 0.93$, $p = .38$. Figure 6 presents funnel plots of the whole sample ($k = 312$) and each literature separately. We quantified visual inspection of plots using Peter's regression test within a three-level model with RVE (Nakagawa & Santos, 2012). The slope of Peter's regression was not significant at the whole sample level, $b = 13.16$, $SE = 12.25$, $p = .29$, 95% CI [−11.92, 38.23], nor when performed separately in the attention set literature, $b = 23.79$, $SE = 15.55$, $p = .14$, 95% CI [−8.81, 56.39], and load literature, $b = -12.67$, $SE = 17.32$, $p = .48$, 95% CI [−50.36, 25.02].

However, publication bias and moderation might exist simultaneously. Because there was significant heterogeneity in our summary estimates, the possibility that bias might impact moderator estimates is also worth considering. To this end, we resubmitted each moderator to separate three-level meta-analytic Peter's regression tests with RVE. Four analyses showed indication of potential bias within the attention set literature: the validity check "online," the inhibition analysis, and the number of potential targets and distractors. The slope for measure of precision in the former was significant, $b = 41.41$, $SE = 15.69$, $p = .02$, 95% CI [7.83, 74.99], as were the latter when using a more conservative p value cutoff ($p < .10$), $b = 23.16$, $SE = 13.07$, $p = .09$, 95% CI [−4.24, 50.56], $b = -88.80$, $SE = 40.55$, $p = .09$, 95% CI [−199.09, 21.49], $b = 26.20$, $SE = 14.57$, $p = .09$, 95% CI [−4.38, 56.78], and $b = 24.26$, $SE = 14.01$, $p = .10$, 95% CI [−5.14, 53.65], respectively. The overall summary estimate also showed potential indication of bias in our sensitivity analysis with cognitive load studies reverse-coded, $p = .06$ (see Supplemental Materials).

We proceeded with sensitivity analyses where summary estimates and moderators were reestimated with a data set trimmed of the low precision cases, and where summary estimates and moderators were

Figure 6

Inverse Sample Size as a Function of Effect Size (Log Odds Ratio) for Total Sample and Each Literature/Theory Separately



Note. Funnel plots of the total sample (left), attention set (middle), and load (right). Effect size (x-axis) reflects log odds ratios (*OR*), which can be interpreted as the odds of noticing the critical stimulus in the treatment group (e.g., attention set match) compared with the control group (e.g., attention set mismatch). A log *OR* of 0 reflects equal rates of noticing between groups. A log *OR* greater than 0 indicates an increase in noticing in the treatment group compared with the control group. Reference lines indicate summary estimate.

reestimated based on Peter's regression, focusing on the model intercept as an estimate of the adjusted effect accounting for study precision. Adjusted summary estimates alongside uncorrected summary estimates for the whole literature and for each theoretical effect separately are presented in Table 5, while results of moderator analyses are presented in Supplemental Materials. Overall, these analyses suggested that results were largely robust to any potential publication bias, although note that the load summary estimate was nonsignificant when using bias adjusted estimates. The worst-case meta-analysis summary estimates are also reported in Table 5 and suggest that, even under a "worst-case" scenario, publication bias could not attenuate the overall summary estimate or the attention set summary estimate to the null, whereas the load summary estimate was rendered nonsignificant under worst-case conditions (Mathur & VanderWeele, 2020).

Discussion

The present meta-analysis sought to quantitatively synthesize research on IB with focus on its two most prominent theories: attention set and load theory. IB rates were found to vary in alignment with predictions made by each theoretical construct. Based upon conventions (Chen et al., 2010), the summary estimate for attention set (OR = 3.26, 95% CI [2.33, 4.57]) approximates a medium effect, while the summary estimate for load (OR = 1.75, 95% CI [1.10, 2.79]) approximates a small-to-medium effect. These summary estimates should prove useful for future work on IB, for example, by informing statistical power calculations to better identify theoretical and methodological relationships.

Table 7 summarizes our key theoretical and methodological findings. Relevant to our aims, the difference between the attention set summary estimate and load summary estimate did not reach statistical significance (p = .08), suggesting that the two theories are comparable in magnitude. However, several caveats to this conclusion need be acknowledged. For one, the difference between summary estimates *was* significant in sensitivity analyses with cognitive load studies reverse-coded (p < .02, see Supplemental Materials). Furthermore, the attention set and load summary estimates each represent a combination of theoretical subconstructs, which one might argue are unreasonable to compare at the aggregate in consideration that theoretical effects were, in some cases, moderated by these distinct subconstructs. On the other hand, if true differences did exist, then one means of circumventing these issues might simply be to compare the most robust subconstruct of each theory; and yet, a post hoc comparison (not reported here) of the strongest variety of each—that is, a

Table 7

Key Theoretical and Methodological Findings

Theoretical findings
1. The effects of attention set and load on IB are comparable in magnitude.
2. An attention set match appears to attenuate the effect of load on IB.
3. Inhibition likely contributes toward the effect of attention set on IB.
4. Explicitly instructed attention sets are stronger than implicit attention sets, however an attention set effect reliably occurs even in the absence of explicit task instructions.
5. A feature-based attention set is the strongest kind of attention set.
6. Perceptual load is a robust and reliable predictor of IB; however, more work is needed to clarify the role of cognitive load in IB.
Methodological findings
1. Paradigm does not influence summary estimates, nor does it confound theoretical effects.
2. Primary task moderates the overall summary estimate.
3. The number of targets and number of distractors moderate attention set.
3. Critical stimulus exposure time (as a proportion of trial duration) moderates attention set.
4. Task setting does not moderate summary estimates, suggesting good external validity to our findings, but there is an overrepresentation of artificial task settings in the literature.
5. Study quality impacts effect size magnitude, and the full attention trial exclusion criterion moderates the load summary estimate.[a]

Note. IB = inattentional blindness.
[a] Full attention trial exclusion criterion did not retain significance across all publication bias estimates (refer to Supplemental Table 5).

feature-based attention set and perceptual load—was *not* significant ($p = .10$). It therefore appears that the two theories may not be as different as many might believe, at least quantitatively speaking.

To gauge the reliability of any conclusion that can be drawn, it is also important to evaluate the overall quality of evidence. To this point, while our results are based on a reasonable number of studies ($K = 81$), it is noteworthy that prediction intervals imply the effect of each respective theory will, within 95% of studies conducted, potentially produce a null effect. This alone may be cause for concern; however, we are hesitant to suggest it threatens the validity of either theory. For one, heterogeneity in meta-analysis is the "norm." Moreover, visual inspection of forest plots indicated that there was general consistency in effect size direction (see Supplemental Figures 2–6). A possible explanation for the heterogeneity may be that variance emerges because both theories are better described as continuous constructs and yet are dichotomized for the purpose of experimentation. If each theory describes a mechanism that exists on a spectrum, it follows that there will exist a set of experimental parameters where attention set and/or load does *not* contribute to rates of IB. Prediction intervals may therefore be capturing true variance in effect sizes based on how effectively each theoretical construct has been manipulated. This implies that heterogeneity emerges due to differences within (the manipulation of) the theoretical construct itself. This is an important consideration in relation to the difference between theories, as one lends itself more favorably to categorical distinctions (attention set) than the other (load) and yet showed greater heterogeneity—a point we expand on below.

How Does the Attention Set Summary Estimate Vary by Type and Level?

At their foundation, each theory can be scrutinized by asking whether it offers explanatory power beyond its own methodological operationalization. Undoubtedly, a good theory should generalize beyond the method used to operationalize it. In this context, the attention set moderator "type" might be the most important theoretical manipulation examined, as it provides indication of the effect of the manipulation (whether the attention set is instructed or not). Our results showed that explicit attention sets were significantly larger ($OR = 4.62$, 95% CI [3.15, 6.78]) than those set implicitly ($OR = 2.00$, 95% CI [1.51, 2.64])—but crucially, implicit attention sets still produced a significant effect. This indicates that, although the instructional manipulation is clearly effective, it is not required for an effect to be observed. Because this moderator reflects a key manipulation employed within the attention set literature, we consider this finding quite favorable for the attention set account.

As predicted, the processing "level" of the attention set (feature, semantic, inherent) was found to moderate the attention set summary estimate. While prior literature suggested categorical (i.e., semantic) attention sets may be stronger than feature-based attention sets (Koivisto et al., 2004), this does not appear to pan out when the literature is synthesized. Attentional control settings for bottom-up stimulus properties were larger than those of inherent attention sets and semantic attention sets (although the latter did not reach statistical significance). We ought to point out that, if quantity of data permitted, the processing levels of attention set could feasibly be subcategorized further. For example, feature-based effect sizes included, among others, those of shape (e.g., Most et al., 2005, Experiment 1) and color (e.g., Drew & Stothart, 2016). Semantic effect sizes included, among others, categorical (e.g., Most, 2013), context-based (e.g., Pammer et al., 2015), and relational sets (e.g., Goldstein & Beck, 2016). Inherent attention sets included the characteristic "threat superiority" and "animate monitoring" hypotheses (e.g., Calvillo & Hawkins, 2016), as well as more abstract sets that hinged on individual differences, such as observers with social anxiety noticing negative affect faces more (Lee, 2009). The processing of such a wide variety of task properties will necessarily be widely distributed and involve complex interactions across the cortex; and hence, it is likely that differences that occurred *within* the attention set categories will have contributed to heterogeneity of the attention set summary estimate.

What Is the Role of Inhibition in the Attention Set Effect?

Within the sample of studies to have examined attentional inhibition, we observed larger differences in noticing between the match versus mismatch conditions, than the match versus neutral conditions. What this suggests is that attention-based enhancement (when the stimulus is a match with the target category) cannot fully account for the rates of noticing that were observed in this subset of studies. However, we must acknowledge several limitations with this analysis. The first is that the distance between conditions could not always be controlled for. For example, in Most et al. (2001, Experiment 2), targets were always gray, distractors were either white or black, and the critical stimulus was either white or black; hence, the critical stimulus was equally "distant" (in units of luminance) from the target in both conditions, but in one condition was a match with distractors (inhibition condition), and the other was not (neutral condition). The neutral condition was therefore a greater "distance" from the distractor than the target, an issue that might confound its use as a neutral condition—dependent on how attentional settings operate.

Later work by Goldstein and Beck (2016) suggested a "relational" set may be established regarding the relationship between targets and distractors. For example, with gray targets and black distractors, a relational set may be for "lighter" objects. In this case, a white critical stimulus would no longer be neutral but would instead constitute a *relational* match. The broader concern this raises is that what constitutes a "neutral" stimulus is contingent on one's theoretical assumption of how attentional settings operate in the first place. Does attention function in a categorical fashion, where an unexpected object is automatically categorized as either a "match" or "mismatch," according to which it most closely matches? If so, it would appear that no true "neutral" stimulus can exist. For our purposes, to avoid circularity, we intended *not* to select neutral conditions with this (or any particular) theoretical postulation in mind, nor with any particular study outcome in mind. However, asymmetries in the comparisons made in this analysis will invariably arise due to the true nature of how attentional settings operate. Nonetheless, in our view, our analyses were a reasonably conservative attempt to assess whether inhibition contributes to the attention set effect, precisely because it relied on testing whether differences in noticing emerged between the "neutral" and "mismatch" conditions, that the attention set theory will make a null prediction on.

How might future studies more precisely examine the effect of inhibition? An intriguing line of work might be to further examine how attention-based enhancement and inhibition interact, such as under conditions where they conflict. For example, in the second experiment of Yan et al. (2012), participants attended to white circles, while ignoring two white and two black distractor circles. Despite the attentional enhancement that would be predicted to occur when the critical stimulus was white, more participants noticed the critical stimulus when it was black than when it was white. Keep in mind that there was no reason whatsoever that the black critical stimulus ought to have been noticed more, as the color black was always a distractor color. The authors interpreted their results as strong evidence for inhibition, reasoning that there was greater need for inhibition of the color white due to the possibility that these distractors could be confused with targets (similar to Koivisto & Revonsuo, 2008).

How Does the Load Summary Estimate Vary by Type?

Turning to load theory, unlike attention set, load faces the criticism that it may not be comparable across studies because the literature lacks a definition of the *difference* between load conditions. One of our aims was to better understand the significance of this issue. When we quantified the difference between load conditions within static search array tasks and subjected this to meta-regression, the "difference score" of load studies did not moderate the load summary estimate (see Supplemental Materials). This is a noteworthy finding: it indicates that, while differential dose effects might be predicted to occur under certain conditions, the small variance in studies observed in this sample did not impact summary estimates. Of course, this cannot be generalized to the load literature more broadly, as it was only evaluated within studies employing a static search array task (e.g., Calvillo & Jackson, 2014). It is an open question whether this finding holds for studies that have used more disparate methods of manipulating load (Zhang et al., 2018). Nevertheless, this subset of studies seems a good starting place for an understanding of the validity of comparing differential load effects across studies. Our findings should not be taken to mean that researchers should shy away from this issue. Rather, they emphasize that the absence of a consistent definition requires addressing. Doing so could help advance the reliability of the load account of IB and may even point to new predictions concerning threshold or load-related tipping points. Possible ideas for future research might be to examine whether load effects are linear or follow some other function, or perhaps whether differences between load conditions are better captured as a ratio of the overall difference (e.g., a difference of 80% between conditions), rather than an absolute difference (i.e., a difference of four items).

In relation to differences between perceptual load and cognitive load, results suggested that the perceptual load and cognitive load summary estimates were comparable (see Table 6). However, both of our load sensitivity analyses suggested a significant difference between the distinct types of load (refer to Supplemental Materials). Most notably, while the perceptual load summary estimate was robust to all sensitivity analyses, the cognitive load summary estimate was rendered nonsignificant within sensitivity analysis using a different coding scheme and where cognitive load effects were reverse-coded. In our view, this issue arises due to conceptual issues with the cognitive load construct itself. Irrespective of how cognitive load is interpreted, there is undoubtedly more inconsistency in cognitive load effects than any other theoretical construct in the IB literature. Ultimately, this set of findings substantiates that perceptual load is a robust theoretical predictor of IB, but that more work is needed to clarify the role of cognitive load in IB. Until this is done, we must conclude that load theory's position with respect to cognitive load has yet to be verified.

How Do Attention Set and Load Interact?

Toward our aim of understanding how attention set and load interact, analysis of the sample of studies that examined both theoretical constructs indicated that the pooled estimate for load was not significant when there was an attention set match. We must exercise some caution with our interpretation of this finding, given that the overall moderating effect of load in this model was not significant ($p = .13$, see Supplemental Materials). Nevertheless, this finding is quite striking—it suggests that differences in noticing between low load and high load are negligible when there is an attention set match. In returning to our point raised in the introduction, to the extent that the attention set and load theories are at odds, for example, in the circumstance that they make conflicting predictions, our results support the prediction made by the attention set account, rather than the load account of IB. In further support of the attention set account, the pooled estimate for attention set remained significant under conditions of *both* high load and low load, and hence, the same effect was not found in reverse (i.e., load did not appear to abolish the attention set effect). In fact, the effect of attention set was larger in magnitude in the high load subgroup (albeit this difference did not reach statistical significance), suggesting that conditions of high load might strengthen the effect of attention set—although this difference will have presumably occurred due to higher rates of IB in the mismatch condition under high load.

Aside from the modest sample ($K = 9$), a key limitation in our interaction analysis was that half of the sample had implemented an "inherent" attention set. In some cases, the notion of an inherent attention set could be subject to circular reasoning. For example, a happy face captures attention more quickly than a sad face, suggesting that attention is inherently set for happy faces. Attention is thought to have been "set" in these studies because the stimulus was hypothesized to be inherently meaningful and would therefore implicitly capture attention (Gupta & Srinivasan, 2015). The issue is that we can shift our reasoning to allocate the labeling of a set match versus mismatch based on whatever happened to capture attention more quickly. Thus, attention can be set in these studies post hoc. Without verifiable a priori hypotheses, there is always the possibility that these findings are the result of hypothesizing after the results are known (i.e., "HARKing"). To be clear, we are by no means suggesting that this has occurred in the included studies. We are, however, highlighting the possibility of this occurring within the attention set literature more broadly. To improve transparency in this body of work, it will be important to place greater emphasis on preregistration of hypotheses in future attention set experiments. We also recommend that future efforts better establish the reliability of this set of findings by placing more emphasis on the interaction between load and feature-based attention sets in IB in future.

Task Parameter Moderators

The second of our principal aims was to address how different design features might introduce bias or otherwise moderate theoretical effects. To this end, we analyzed a range of common task parameters for their moderating effects on summary estimates. The overall conclusions that can be drawn from these analyses are as follows: (a) neither paradigm nor setting moderate summary estimates, however, the primary task does impact effect sizes, (b) critical stimulus exposure time (as a proportion of trial duration) and the number of objects within the visual display are important factors in determining the strength of the attention set effect; and the latter might provide a previously underutilized task parameter to test each theory against one another, and (c) study quality appears to moderate summary estimates, in particular with respect to the load summary estimate and whether participants were considered for exclusion based on the full attention trial. We expand on these findings below.

Paradigm and Stimulus Exposure Time

Contrary to our suspicions, paradigm did not moderate the overall summary estimate, nor was there any evidence of this task parameter biasing theoretical moderators (see Supplemental Materials). This implies that methodological choices concerning the nature of the paradigm employed (whether the visual display had motion or remained stationary) do not attenuate theoretical effects—an optimistic finding that suggests that neither theory is reducible to differences arising based on whether the stimulus display contained motion (Kreitz, Furley, et al., 2015; Most, 2010). Similarly, whether the study used a real-world (i.e., naturalistic) or computer-based (i.e., artificial) setting did not impact effect sizes. This finding corroborates the generalizability of these findings to real-world occurrences of IB, because it suggests that differences in theoretical effects are equivocal whether within a computer-based setting or real-world context—though it need be cautioned that natural settings were used relatively infrequently in our sample.

With respect to the attention set theory, attention set effect sizes tended to be larger with a greater critical stimulus exposure time (as a proportion of trial duration). The significance of this moderator appeared to suggest that at least some of the difference between the attention set and load summary estimates is due to stimulus exposure time. This might add weight to the conclusion that there is no true difference in the strength of effect between attention set and load. We are hesitant to draw this conclusion, given that only the attention set summary estimate was impacted by the critical stimulus exposure time proportion and because the effect did not appear to be driven by differences in the mere timing of stimuli (as stimulus duration was overall comparable between theories, see Table 4). The effect also did not appear to be driven by timing differences between paradigms, as it remained even after accounting for paradigm (p = .02, not reported here).

Another possible interpretation is that there may be theoretical reasons why an effect would be expected to occur for one theory and not the other. For example, processing requirements of any given task presumably remain consistent throughout the duration of an experimental trial, and hence, load-related effects would not be expected to change over the duration of an experiment. Conversely, all else being equal, an attention set may "strengthen" over an experimental trial as attention becomes increasingly tuned to the task. One of the only studies that experimentally examined this moderator comes from Kreitz, Furley, and Memmert (2016; but also see White & Davies, 2008; Beanland & Pammer, 2010). These authors reported a risk ratio of 1.58 in relation to noticing the critical stimulus with greater exposure time (4 s) compared with less exposure time (1.8 s; Experiment 1); however, they did not manipulate either theoretical construct. It would be worthwhile to test whether attention set and load effects are comparable by manipulating both theoretical constructs within a similar task while simultaneously manipulating exposure time (as a proportion of the trial duration) between different theoretical conditions.

Primary Task, the Number of Targets, and the Number of Distractors

Results indicated that effect sizes were larger in magnitude for search-based tasks compared with discrimination tasks (see Table 5). Interestingly, effect sizes within count-based tasks were also larger than discrimination tasks; however, this difference did not survive Bonferroni correction (p = .15, corrected). Note that, while the interaction was not significant, the effect of primary task was largely driven by the attention set summary estimate (p = .03 vs. p = .21, when examined in separate models of each respective summary estimate). This is interesting because, in terms of task requirements, possibly the most obvious difference is that both count and search tasks require varying degrees of search-based attentional selection operating toward several objects, such as maintaining the count of several targets in the former or an explicit visual search in the latter. Comparatively, discrimination tasks characteristically require attention to a single object, with high demands placed on visual acuity within, or about, that object (for classic examples of each, see Figure 2). Notably, many of the mechanisms that govern attention in the former, such as those operating during visual search, can be readily subsumed underneath the attentional settings of the observer (for review, see Wolfe & Horowitz, 2017). Hence, it is perhaps no wonder that effect sizes tend to be stronger in these tasks.

Of further interest, when compared with discrimination tasks, both count and search-based tasks tended to have more targets (both $p < .001$) and distractors ($p < .001$ for the former, p = .09 for the latter), and both of these variables also moderated the attention set summary estimate in the same direction, such that attention set effects were larger with an increase in the number of targets and/or distractors (note that the latter did not retain significance across all publication bias analyses, see Supplemental Materials). Thus, the mere number of objects may, in fact, be a key moderator of the attention set effect. But because an increase in the number of objects will also increase load, this begs several important questions—namely (a) are these effects "merely" the result of load or are the theoretical effects of load instead better "explained away" as a product of task parameters and (b) can the number of objects offer a means to test both theoretical constructs against one another?

In addressing these questions, we wish to raise several points for consideration. First, post hoc analysis (not reported here) where we removed all studies that correspondingly varied load alongside attention set no longer showed a significant effect for the number of targets (p = .07). This suggests that the moderating effect of the number of targets may be at least partly attributable to load. Second, the effect of load is observed across a variety of manipulations, not merely those where the number of objects is varied. Load theory's construct validity is therefore probably robust (even if the finer details of the theory require clarification). Third, a follow-up experiment could seek to clarify whether the effect of load and the number of objects can be dissociated. A clever illustration for how this might be achieved comes from Wright et al. (2013). In their study, under conditions of low load, participants tracked a single letter among seven distractor letters, whereas under conditions of high load, participants tracked four letters among four distractor letters. Hence, while load varied based upon how many objects the participants had to explicitly track, the number of objects were equated between conditions of high load and low load.

The fact that attention set effect sizes are larger with an increase in the number of targets highlights the possibility that this reflects a novel task parameter to test both theories against one another. Load theory would predict that systematically increasing the number of targets should increase load and thus lead to a *decrease* in rates of noticing (assuming a perceptual load manipulation). However, based on our findings, increasing the number of targets should produce larger attention set effects, and therefore increase rates of noticing (or at the very least remain comparable) when the critical stimulus is an attention set match. Because this reflects an easily modifiable task parameter, it would be relatively simple to quantify the precise tipping point between each theory, for example, the point where load "wins out" against attention set.

Study Quality and the Full-Attention Trial Exclusion Criterion

The validity check concerning the full-attention trial exclusion criterion was a significant moderator of the load summary estimate. We must emphasize that the criterion was used in the present meta-analysis primarily as a check for study quality. This meant that a study was catalogued as having met the criterion if there was an indication that it had *considered* excluding participants based on the criterion, rather than necessarily having done so. It is therefore not entirely correct to suggest that excluding participants based on the full-attention trial impacts effect sizes per se. For a more in-depth discussion of the issues surrounding the full-attention trial exclusion criterion, the reader is referred to White et al. (2018). Their overarching point is that whether the exclusion is employed can, in some cases, render a null result significant (and in other cases, a significant result null). The present findings certainly emphasize that whether studies satisfied this criterion impacts effect sizes, where slightly larger effects are found in experiments that made mention of the exclusion criterion.

In our view, there are two nonmutually exclusive possibilities for how the moderating effect of the full attention trial exclusion criterion can be interpreted: (a) explicit mention of the criterion by study authors is a marker of study quality, and therefore study quality impacts effects sizes or (b) the direct use of the criterion impacts effect sizes, such as through the exclusion of poor-quality data (e.g., excluding participants who were not performing the task properly). We are inclined to surmise the former, though not at the exclusion of the latter. This is because our overall quality score measure *did* indicate that studies that scored a zero out of three (those that did not meet any of our validity check criteria) had smaller effect sizes compared with those that scored a three out of three (those that met all of our validity check criteria). Consequently, there is reason to suspect study quality does impact effect sizes in the IB literature more broadly. Arguably, this might relate to our previous point regarding the dichotomization of each theory—higher quality studies may yield larger effects because these researchers have effectively maximized the difference between theoretical conditions. Ultimately, the significance of this moderator is intriguing and emphasizes that the choice of whether to employ the criterion is not a trivial one. We recommend that, whether authors choose to adopt the full-attention trial exclusion criterion or not, the choice be explicitly outlined with justification given to the reasons why it was not used, if applicable.

Limitations

Beyond those already mentioned, what additional components might explain heterogeneity in summary estimates that were not considered in the present meta-analysis? First, given our principal aims were geared toward theoretical and methodological clarification, we did not strongly consider how characteristics of the sample might have affected the results. Previous studies have found mixed evidence for age and sex differences in IB (Hannon & Richards, 2010; Kreitz, Schnuerch, et al., 2015; Richards et al., 2010). It must be stressed that even basic demographic information such as age, sex, and ethnicity was missing from many studies sampled here (particularly those published more than 10 years ago in the load literature). For example, less than five studies in our sample detailed ethnicity data of their sample. A potentially more pressing concern is whether the study opted for a convenience sample (e.g., undergraduate students) or community-based sample. While this was a methodological choice not considered here, this is an important factor worth future investigation in consideration that convenience samples are not likely to be representative of the general population (Jager et al., 2017).

Limitations with our sample need also be acknowledged. At the onset, we suspected there would be an approximately equal distribution of attention set and load studies in the IB literature. However, there is a clear bias toward attention set studies (and its sub theories). This suggests that a more reliable assessment and comparison of theories might be possible in future, should more studies on load be conducted. In addition, despite our comprehensive search strategy, we must acknowledge the possibility that studies may have been missed and, if these were to be included, the pattern of results may vary from those reported here. Finally, while we included our own study quality checks, to provide a more thorough and valid picture of the quality of data available, we may have benefitted from using a verified study quality assessment tool (Ma et al., 2020). Indeed, the fact that our results suggest studies with higher quality scores showed larger effect sizes reenforces this point.

An important variable worthy of future consideration is the selection criteria for demarcating a participant as a "noticer" or as "inattentionally blind." This was a methodological feature initially considered for extraction. However, it became burdensome due to clarity in the reporting of this feature in studies. This is important because, to establish a reliable research program on IB, studies ought to first agree upon and outline the kind of evidence considered "sufficient" for any given participant to be considered a noticer. Is the use of forced-choice questioning reasonable given significant rates of correct guessing in those considered inattentionally blind (Kreitz et al., 2020)? The earlier work of Simons and Chabris (1999) adopted so-called "funnel" questioning, where participants were probed using increasingly specific questions—though this method appears to have fallen out of favor in recent years. Overall, these are issues that have, to our understanding, not been given adequate discussion in the IB literature. Our efforts to avert these issues is evidenced by the fact that we made no prior assumptions for what was required as an awareness outcome measure. Yet, these are foundational issues that impact upon how IB is studied; therefore, it is essential that these are clarified by the IB research community more broadly.

This allows us to reflect on our own definition of IB. We began by reasoning that the common factor across experimental occurrences of IB is that it occurs when a stimulus is unexpected. Yet, it is undoubtedly common in real-world settings that a failure to see something directly within one's field of view can occur, even when expecting its presence to some extent (such as the driver who looked but failed to see a pedestrian). Hence, "expectedness" is itself probably not a dichotomous construct, but rather will exist in degrees of magnitude. Given that our definition of IB hinges on a manipulation of expectancy, expectation itself ought to be more directly considered an experimental variable in future (such as in Horstmann & Ansorge, 2016).

Implications, Recommendations, and Conclusion

What are the broader implications of our findings for the psychologist, neuroscientist, and practitioner? For the psychologist, our findings substantiate that an observer's attention set has the capacity to provide a stimulus with processing capabilities largely unaffected by load. This conflicts with the first assumption of load theory, as it appears to reflect perceptual processing of unexpected information despite capacity exhaustion ("late" attentional selection despite high load). Load theory faces the challenge of rectifying this pattern of findings and future efforts ought to seek to clarify how attentional settings can afford processing benefits despite load under conditions of IB. For the neuroscientist, these findings may help in better understanding the nature of conscious processing—namely, that inhibition likely contributes to IB (Wood & Simons, 2017b) and that unexpected objects can be preferentially boosted into conscious awareness based purely on their semantic properties (Koivisto & Revonsuo, 2009), both imply a considerable depth of "unconscious" processing must occur under conditions of IB—a similar conclusion to that reached by both Kreitz et al. (2020) and Nobre et al. (2020). The next question, then, is where or when within the visual processing cortical hierarchy does this unconscious processing bottleneck cease? Finally, for the practitioner, findings can—with some caution—be generalized to real-world occurrences of IB. For example, because attention set effects are stronger under high load, it might (counterintuitively) be useful to place more demand on attention to improve task focus under certain scenarios. On the other hand, in circumstances where IB holds potentially dire consequences, such as in the case of a surgeon missing a tumor during surgery, it might benefit to simply prepare in advance—even something as simple as the potential colors or shapes that a tumor may take—as findings suggest that an attention set for such features should strongly decrease rates of IB, even under stressful and high demand conditions.

Before concluding, we must stress several recommendations for the benefit of scientific rigor in this area moving forward. First, it is crucial that future work adheres to basic demographic reporting standards. Age, gender, and ethnicity data need be reported in studies—across all experiments and conditions, if possible. Second, all hypotheses and study design choices should, ideally, be preregistered or use a registered protocol. This includes any directional hypotheses, the rationale for how hypotheses will be tested, and the decision criteria for demarcating a participant as a "noticer" or as "inattentionally blind" (for a good example, see Wood & Simons, 2017a). If this is not possible, authors ought to state their hypotheses up front, transparent to the process by which they were derived, and explicitly highlight if they deviated from their decision in how participants will be coded (and the reason why, if appropriate). Third, explicit justification should be given to the choice of all treatment and control groups, and considerations for exclusionary criteria need be made clear—were participants questioned on their prior knowledge of IB, and were these participants considered for exclusion? Were participants considered for exclusion based upon their performance on the full-attention trial or on their overall task performance? Finally, in relation to theoretical manipulations, it is crucial that any theoretical manipulation employed is independently verified using a method that is not confounded with the manipulation itself.

Through addressing several theoretical and methodological issues that have persisted across two bodies of literature on IB, we hope to have advanced the field by providing unique insights that will stimulate future research in this area. Overall, to better ensure scientific rigor and reliability moving forward, efforts need be placed on advancing a more coherent and unified research program in the IB literature, one where the strengths of the varying theories and methodologies can be exploited, and their biases can be succinctly cross-examined and scrutinized.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Aimola Davies, A. M., Waterman, S., White, R. C., & Davies, M. (2013). When you fail to see what you were told to look for: Inattentional blindness and task instructions. *Consciousness and Cognition*, *22*(1), 221–230. https://doi.org/10.1016/j.concog.2012.11.015

Al-Moteri, M. O., Symmons, M., Cooper, S., & Plummer, V. (2018). Inattentional blindness and pattern-matching failure: The case of failure to recognize clinical cues. *Applied Ergonomics*, *73*, 174–182. https://doi.org/10.1016/j.apergo.2018.07.001

Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. (2006). Multiple-target tracking: A role for working memory? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *59*(6), 1101–1116. https://doi.org/10.1080/02724980543000097

Almalik, O., Zhan, Z., & van den Heuvel, E. R. (2021). Tests for publication bias are unreliable in case of heteroscedasticity. *Contemporary Clinical Trials Communications*, *22*, Article 100781. https://doi.org/10.1016/j.conctc.2021.100781

Assink, M., & Wibbelink, C. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

*Barnhart, A. S., & Goldinger, S. D. (2014). Blinded by magic: Eye-movements reveal the misdirection of attention. *Frontiers in Psychology*, *5*, Article 1461. https://doi.org/10.3389/fpsyg.2014.01461

*Beanland, V., Allen, R. A., & Pammer, K. (2011). Attending to music decreases inattentional blindness. *Consciousness and Cognition*, *20*(4), 1282–1292. https://doi.org/10.1016/j.concog.2011.04.009

*Beanland, V., & Chan, E. H. (2016). The relationship between sustained inattentional blindness and working memory capacity. *Attention, Perception & Psychophysics*, *78*(3), 808–817. https://doi.org/10.3758/s13414-015-1027-x

*Beanland, V., & Pammer, K. (2010). Looking without seeing or seeing without looking? Eye movements in sustained inattentional blindness. *Vision Research*, *50*(10), 977–988. https://doi.org/10.1016/j.visres.2010.02.024
Beanland, V., Tan, C. H., & Christensen, B. K. (2018). The unexpected killer: Effects of stimulus threat and negative affectivity on inattentional blindness. *Cognition and Emotion*, *32*(6), 1374–1381. https://doi.org/10.1080/02699931.2017.1394269
*Becker, M. W., & Leinenger, M. (2011). Attentional selection is biased toward mood-congruent stimuli. *Emotion*, *11*(5), 1248–1254. https://doi.org/10.1037/a0023524
*Brailsford, R., Catherwood, D., Tyson, P. J., & Edgar, G. (2014). Noticing spiders on the left: Evidence on attentional bias and spider fear in the inattentional blindness paradigm. *Laterality: Asymmetries of Body, Brain, and Cognition*, *19*(2), 201–218. https://doi.org/10.1080/1357650X.2013.791306
*Calvillo, D. P., & Hawkins, W. C. (2016). Animate objects are detected more frequently than inanimate objects in inattentional blindness tasks independently of threat. *The Journal of General Psychology*, *143*(2), 101–115. https://doi.org/10.1080/00221309.2016.1163249
*Calvillo, D. P., & Jackson, R. E. (2014). Animacy, perceptual load, and inattentional blindness. *Psychonomic Bulletin & Review*, *21*(3), 670–675. https://doi.org/10.3758/s13423-013-0543-8
*Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattentional blindness. *Cognition*, *102*(3), 321–340. https://doi.org/10.1016/j.cognition.2006.01.002
*Chabris, C. F., Weinberger, A., Fontaine, M., & Simons, D. J. (2011). You do not talk about Fight Club if you do not notice Fight Club: Inattentional blindness for a simulated real-world assault. *i-Perception*, *2*(2), 150–153. https://doi.org/10.1068/i0436
*Chai, Y. (2004). *The effects of unexpected luminance and color changes on the perception of a single complex object* [Unpublished doctoral dissertation]. University of Texas Arlington.
Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation*, *39*(4), 860–864. https://doi.org/10.1080/03610911003650383
*Chen, H., Wang, Z., & Wang, L. (2011). *Attention set: The effects of consistency of stimulus type and number expectation on inattentional blindness* [Poster presentation]. European Conference on Visual Perception, Toulouse, France. https://f1000research.com/posters/1089366
Cheng, K., Yang, K., Qin, L., Zhuo, Y., & Yan, H. (2019). Perceptual load modulates contour integration in conscious and unconscious states. *PeerJ*, *7*, Article e7550. https://doi.org/10.7717/peerj.7550
Cheung, M. W. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968
Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427
*Clement, A., Stothart, C., Drew, T., & Brockmole, J. R. (2019). Semantic associations do not modulate the visual awareness of objects. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *72*(5), 1224–1232. https://doi.org/10.1177/1747021818811411
Couperus, J. W., & Lydic, K. O. (2019). Attentional set and the gradient of visual spatial attention. *Neuroscience Letters*, *712*, Article 134495. https://doi.org/10.1016/j.neulet.2019.134495
*de Fockert, J. W., & Bremner, A. J. (2011). Release of inattentional blindness by high working memory load: Elucidating the relationship between working memory and selective attention. *Cognition*, *121*(3), 400–408. https://doi.org/10.1016/j.cognition.2011.08.016
Deeks, J. J., Macaskill, P., & Irwig, L. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology*, *58*(9), 882–893. https://doi.org/10.1016/j.jclinepi.2005.01.016
*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*(4), 552–568. https://doi.org/10.1068/p6049
*Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition*, *93*(1), B27–B38. https://doi.org/10.1016/j.cognition.2003.10.010
*Drew, T., & Stothart, C. (2016). Clarifying the role of target similarity, task relevance and feature-based suppression during sustained inattentional blindness. *Journal of Vision*, *16*(15), Article 13. https://doi.org/10.1167/16.15.13
*Ersin, K., Gundogdu, O., Kaya, S. N., Aykiri, D., & Serbetcioglu, M. B. (2021). Investigation of the effects of auditory and visual stimuli on attention. *Heliyon*, *7*(7), Article e07567. https://doi.org/10.1016/j.heliyon.2021.e07567
*Fick, C. S. (2007). *Flying under the radar: Studying inattentional blindness in a dynamic task* [Unpublished doctoral dissertation]. Rice University. https://scholarship.rice.edu/handle/1911/20604
Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *robumeta: Robust variance meta-regression* (R package Version 2.0) [Computer software]. https://CRAN.R-project.org/package=robumeta
Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030
*Fougnie, D., & Marois, R. (2007). Executive working memory load induces inattentional blindness. *Psychonomic Bulletin & Review*, *14*(1), 142–147. https://doi.org/10.3758/BF03194041
*Furley, P., Memmert, D., & Heller, C. (2010). The dark side of visual awareness in sport: Inattentional blindness in a real-world basketball task. *Attention, Perception & Psychophysics*, *72*(5), 1327–1337. https://doi.org/10.3758/APP.72.5.1327
*Gao, H., & Jia, Z. (2017). Detection of threats under inattentional blindness and perceptual load. *Current Psychology*, *36*(4), 733–739. https://doi.org/10.1007/s12144-016-9460-0
Gibbs, R., Davies, G., & Chou, S. (2016). A systematic review on factors affecting the likelihood of change blindness. *Crime Psychology Review*, *2*(1), 1–21. https://doi.org/10.1080/23744006.2016.1228799
*Goldstein, R. R., & Beck, M. R. (2016). Inattentional blindness: A combination of a relational set and a feature inhibition set? *Attention, Perception & Psychophysics*, *78*(5), 1245–1254. https://doi.org/10.3758/s13414-016-1091-x
*Graham, E. R., & Burke, D. M. (2011). Aging increases inattentional blindness to the gorilla in our midst. *Psychology and Aging*, *26*(1), 162–166. https://doi.org/10.1037/a0020647
*Greene, C. M., Murphy, G., & Januszewski, J. (2017). Under high perceptual load, observers look but do not see. *Applied Cognitive Psychology*, *31*(4), 431–437. https://doi.org/10.1002/acp.3335
Greig, P. R., Higham, H., & Nobre, A. C. (2014). Failure to perceive clinical events: An under-recognised source of error. *Resuscitation*, *85*(7), 952–956. https://doi.org/10.1016/j.resuscitation.2014.03.316
*Guo, Y., You, X., & Li, Y. (2016). The role of stimulus type and semantic category-level attentional set in sustained inattentional blindness. *The Japanese Psychological Research*, *58*(4), 332–341. https://doi.org/10.1111/jpr.12127
*Gupta, R., & Srinivasan, N. (2015). Only irrelevant sad but not happy faces are inhibited under high perceptual load. *Cognition and Emotion*, *29*(4), 747–754. https://doi.org/10.1080/02699931.2014.933735
Hannon, E. M., & Richards, A. (2010). Is inattentional blindness related to individual differences in visual working memory capacity or executive control functioning? *Perception*, *39*(3), 309–319. https://doi.org/10.1068/p6379
*Harms, I., Dijksterhuis, C., Jelijs, B., Waard, D., & Brookhuis, K. (2019). Don't shoot the messenger: Traffic-irrelevant messages on variable

message signs (VMSs) might not interfere with traffic management. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 564–575. https://doi.org/10.1016/j.trf.2018.09.011

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2019). *Doing meta-analysis in R: A hands-on guide*. Zenodo. https://doi.org/10.5281/zenodo.2551803

*Harvey, A. J., Bayless, S. J., & Hyams, G. (2018). Alcohol increases inattentional blindness when cognitive resources are not consumed by ongoing task demands. *Psychopharmacology*, *235*(1), 309–315. https://doi.org/10.1007/s00213-017-4772-9

Hopf, J. M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(4), 1053–1058. https://doi.org/10.1073/pnas.0507746103

*Horstmann, G., & Ansorge, U. (2016). Surprise capture and inattentional blindness. *Cognition*, *157*, 237–249. https://doi.org/10.1016/j.cognition.2016.09.005

*Horwood, S., & Beanland, V. (2016). Inattentional blindness in older adults: Effects of attentional set and to-be-ignored distractors. *Attention, Perception & Psychophysics*, *78*(3), 818–828. https://doi.org/10.3758/s13414-015-1057-4

*Hughes-Hallett, A., Mayer, E. K., Marcus, H. J., Pratt, P., Mason, S., Darzi, A. W., & Vale, J. A. (2015). Inattention blindness in surgery. *Surgical Endoscopy*, *29*(11), 3184–3189. https://doi.org/10.1007/s00464-014-4051-3

Hutchinson, B. T. (2019). Toward a theory of consciousness: A review of the neural correlates of inattentional blindness. *Neuroscience and Biobehavioral Reviews*, *104*, 87–99. https://doi.org/10.1016/j.neubiorev.2019.06.003

Hutchinson, B. T., & Bandara, K. H. (2022). *IB meta-analysis 1: Theory and methods*. https://osf.io/nbvqx/?view_only=3d1f5759a4374158b1eccf09e29b0286

Hutchinson, B. T., Bandara, K. H., McGovern, H. T., & Talipski, L. A. (2022). Insights on overflow from failure to report tasks. *Behavioural Brain Research*, *417*, Article 113610. https://doi.org/10.1016/j.bbr.2021.113610

Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). More than just convenient: The scientific merits of homogenous convenience samples. *Monographs of the Society for Research in Child Development*, *82*(2), 13–30. https://doi.org/10.1111/mono.12296

*Jenkins, R., Lavie, N., & Driver, J. (2005). Recognition memory for distractor faces depends on attentional load at exposure. *Psychonomic Bulletin & Review*, *12*(2), 314–320. https://doi.org/10.3758/BF03196378

Jones, A., & Johnstone, M. J. (2017). Inattentional blindness and failures to rescue the deteriorating patient in critical care, emergency and perioperative settings: Four case scenarios. *Australian Critical Care*, *30*(4), 219–223. https://doi.org/10.1016/j.aucc.2016.09.005

*Karns-Wright, T., & Rivardo, M. (2010). Noticing of an unexpected event is affected by attentional set for expected action. *North American Journal of Psychology*, *12*(3), 637–649. https://www.researchgate.net/publication/235978425_Noticing_of_an_Unexpected_Event_is_Affected_by_Attentional_Set_for_Expected_Action

Kennedy, K. D., Stephens, C. L., Williams, R. A., & Schutte, P. C. (2017). Repeated induction of inattentional blindness in a simulated aviation environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1959–1963. https://doi.org/10.1177/1541931213601969

Koivisto, M., Grassini, S., Salminen-Vaparanta, N., & Revonsuo, A. (2017). Different electrophysiological correlates of visual awareness for detection and identification. *Journal of Cognitive Neuroscience*, *29*(9), 1621–1631. https://doi.org/10.1162/jocn_a_01149

*Koivisto, M., Hyönä, J., & Revonsuo, A. (2004). The effects of eye movements, spatial attention, and stimulus features on inattentional blindness. *Vision Research*, *44*(27), 3211–3221. https://doi.org/10.1016/j.visres.2004.07.026

Koivisto, M., Kainulainen, P., & Revonsuo, A. (2009). The relationship between awareness and attention: Evidence from ERP responses. *Neuropsychologia*, *47*(13), 2891–2899. https://doi.org/10.1016/j.neuropsychologia.2009.06.016

*Koivisto, M., & Revonsuo, A. (2007). How meaning shapes seeing. *Psychological Science*, *18*(10), 845–849. https://doi.org/10.1111/j.1467-9280.2007.01989.x

*Koivisto, M., & Revonsuo, A. (2008). The role of unattended distractors in sustained inattentional blindness. *Psychological Research*, *72*(1), 39–48. https://doi.org/10.1007/s00426-006-0072-4

*Koivisto, M., & Revonsuo, A. (2009). The effects of perceptual load on semantic processing under inattention. *Psychonomic Bulletin & Review*, *16*(5), 864–868. https://doi.org/10.3758/PBR.16.5.864

Korevaar, D. A., Salameh, J. P., Vali, Y., Cohen, J. F., McInnes, M., Spijker, R., & Bossuyt, P. M. (2020). Searching practices and inclusion of unpublished studies in systematic reviews of diagnostic accuracy. *Research Synthesis Methods*, *11*(3), 343–353. https://doi.org/10.1002/jrsm.1389

*Kowai-Bell, N. (2005). *Inattentional blindness, change blindness and race/ethnicity* [Unpublished doctoral dissertation]. University of California Santa Barbara.

Kreitz, C., Furley, P., & Memmert, D. (2016). Inattentional blindness is influenced by exposure time not motion speed. *Quarterly Journal of Experimental Psychology*, *69*(3), 495–505. https://doi.org/10.1080/17470218.2015.1055771

Kreitz, C., Furley, P., Memmert, D., & Simons, D. J. (2015). Inattentional blindness and individual differences in cognitive abilities. *PLOS ONE*, *10*(8), Article e0134675. https://doi.org/10.1371/journal.pone.0134675

*Kreitz, C., Furley, P., Memmert, D., & Simons, D. J. (2016). The influence of attention set, working memory capacity, and expectations on inattentional blindness. *Perception*, *45*(4), 386–399. https://doi.org/10.1177/0301006615614465

Kreitz, C., Pugnaghi, G., & Memmert, D. (2020). Guessing right: Preconscious processing in inattentional blindness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *73*(7), 1055–1065. https://doi.org/10.1177/1747021820911324

*Kreitz, C., Schnuerch, R., Furley, P. A., Gibbons, H., & Memmert, D. (2015). Does semantic preactivation reduce inattentional blindness? *Attention, Perception & Psychophysics*, *77*(3), 759–767. https://doi.org/10.3758/s13414-014-0819-8

Kreitz, C., Schnuerch, R., Gibbons, H., & Memmert, D. (2015). Some see it, some don't: Exploring the relation between inattentional blindness and personality factors. *PLOS ONE*, *10*(5), Article e0128158. https://doi.org/10.1371/journal.pone.0128158

*Kuvaldina, M., & Adamyan, N. (2011). To expect means not to expect: Role of anticipation rigidity in inattentional blindness. *Journal of Russian & East European Psychology*, *49*(5), 62–77. https://doi.org/10.2753/RPO1061-0405490504

Lapierre, M. D., Cropper, S. J., & Howe, P. D. L. (2017). Shared processing in multiple object tracking and visual working memory in the absence of response order and task order confounds. *PLOS ONE*, *12*(4), Article e0175736. https://doi.org/10.1371/journal.pone.0175736

*Lathrop, W. B., Bridgeman, B., & Tseng, P. (2011). Perception in the absence of attention: Perceptual processing in the Roelofs effect during inattentional blindness. *Perception*, *40*(9), 1104–1119. https://doi.org/10.1068/p6859

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. https://doi.org/10.1016/j.tics.2004.12.004

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*(3), 143–148. https://doi.org/10.1177/0963721410370295

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: Attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *369*(1641), 1–10. https://doi.org/10.1098/rstb.2013.0205

*Lee, H. J. (2009). *Attentional biases in social anxiety: An investigation using the inattentional blindness paradigm* [Unpublished doctoral dissertation]. University of Texas Austin.

*Légal, J. B., Chekroun, P., Coiffard, V., & Gabarrot, F. (2017). Beware of the gorilla: Effect of goal priming on inattentional blindness. *Consciousness and Cognition*, *55*, 165–171. https://doi.org/10.1016/j.concog.2017.08.004

*Li, J., & Su-Ling, Y. (2007). New objects do not capture attention without a top-down setting: Evidence from an inattentional blindness task. *Visual Cognition*, *15*(6), 661–684. https://doi.org/10.1080/13506280600926695

*Li, X., Tao, Q., Fang, Y., Cheng, C., Hao, Y., Qi, J., Li, Y., Zhang, W., Wang, Y., & Zhang, X. (2015). Reward sensitivity predicts ice cream-related attentional bias assessed by inattentional blindness. *Appetite*, *89*, 258–264. https://doi.org/10.1016/j.appet.2015.02.010

Liao, C., & Chiang, T. (2016). Reducing occupational injuries attributed to inattentional blindness in the construction industry. *Safety Science*, *89*, 129–137. https://doi.org/10.1016/j.ssci.2016.06.010

*Lin, S. H., & Yeh, Y. Y. (2014). Attentional load and the consciousness of one's own name. *Consciousness and Cognition*, *26*(1), 197–203. https://doi.org/10.1016/j.concog.2014.03.008

Ma, L. L., Wang, Y. Y., Yang, Z. H., Huang, D., Weng, H., & Zeng, X. T. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: What are they and which is better? *Military Medical Research*, *7*(1), Article 7. https://doi.org/10.1186/s40779-020-00238-8

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, *20*(4), 641–654. https://doi.org/10.1002/sim.698

Mack, A., & Rock, I. (1998). *Inattentional blindness*. MIT Press. https://doi.org/10.7551/mitpress/3707.001.0001

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Applied Statistics*, *69*(5), 1091–1119. https://doi.org/10.1111/rssc.12440

*Matsuyoshi, D., Ikeda, T., Sawamoto, N., Kakigi, R., Fukuyama, H., & Osaka, N. (2010). Task-irrelevant memory load induces inattentional blindness without temporo-parietal suppression. *Neuropsychologia*, *48*(10), 3094–3101. https://doi.org/10.1016/j.neuropsychologia.2010.06.021

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moreno, S. G., Sutton, A. J., Ades, A. E., Stanley, T. D., Abrams, K. R., Peters, J. L., & Cooper, N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology*, *9*, Article 2. https://doi.org/10.1186/1471-2288-9-2

Most, S. B. (2010). What's "inattentional" about inattentional blindness? *Consciousness and Cognition*, *19*(4), 1102–1104. https://doi.org/10.1016/j.concog.2010.01.011

*Most, S. B. (2013). Setting sights higher: Category-level attentional set modulates sustained inattentional blindness. *Psychological Research*, *77*(2), 139–146. https://doi.org/10.1007/s00426-011-0379-7

*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*(1), 217–242. https://doi.org/10.1037/0033-295X.112.1.217

*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*(1), 9–17. https://doi.org/10.1111/1467-9280.00303

Murphy, G., & Greene, C. M. (2016). Perceptual load induces inattentional blindness in drivers. *Applied Cognitive Psychology*, *30*(3), 479–483. https://doi.org/10.1002/acp.3216

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory—Where are we now, and where should we go next? *Psychonomic Bulletin & Review*, *23*(5), 1316–1340. https://doi.org/10.3758/s13423-015-0982-5

Nakagawa, S., & Santos, E. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, *26*(5), 1253–1274. https://doi.org/10.1007/s10682-012-9555-5

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research*, *51*(13), 1526–1537. https://doi.org/10.1016/j.visres.2010.09.003

*Näsholm, E., Rohlfing, S., & Sauer, J. D. (2014). Pirate stealth or inattentional blindness? The effects of target relevance and sustained attention on security monitoring for experienced and naïve operators. *PLOS ONE*, *9*(1), Article e86157. https://doi.org/10.1371/journal.pone.0086157

*New, J. J., & German, T. C. (2015). Spiders at the cocktail party: An ancestral threat that surmounts inattentional blindness. *Evolution and Human Behavior*, *36*(3), 165–173. https://doi.org/10.1016/j.evolhumbehav.2014.08.004

Nobre, A. 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. https://doi.org/10.1016/j.neubiorev.2020.10.005

*Oktay, B., & Cangöz, B. (2018). I thought I saw "Zorro": An inattentional blindness study. *Nöro Psikiyatri Arşivi*, *55*(1), 59–66. https://doi.org/10.29399/npa.19227

*Palmer, D. B., Yamani, Y., Bobrow, T. L., Karpinsky, N. D., & Krusienski, D. J. (2018). Transient signals and inattentional blindness in a multi-object tracking task. *i-Perception*, *9*(1), 1–9. https://doi.org/10.1177/2041669518754595

*Pammer, K., Bairnsfather, J., Burns, J., & Hellsing, A. (2015). Not all hazards are created equal: The significance of hazards in inattentional blindness for static driving scenes. *Applied Cognitive Psychology*, *29*(5), 782–788. https://doi.org/10.1002/acp.3153

*Pammer, K., & Blink, C. (2013). Attentional differences in driving judgments for country and city scenes: Semantic congruency in inattentional blindness. *Accident Analysis and Prevention*, *50*, 955–963. https://doi.org/10.1016/j.aap.2012.07.026

*Pammer, K., Sabadas, S., & Lentern, S. (2018). Allocating attention to detect motorcycles: The role of inattentional blindness. *Human Factors*, *60*(1), 5–19. https://doi.org/10.1177/0018720817733901

Paparella, S. (2013). How did I miss that? The safety challenges of inattentional blindness. *Journal of Emergency Nursing*, *39*(4), 358–359. https://doi.org/10.1016/j.jen.2013.03.005

*Pazzona, R., Guicciardi, M., & Murgia, M. (2018). The inattentional blindness in soccer referees. *Medicina dello Sport*, *71*(2), 2016–2225. https://doi.org/10.23736/S0025-7826.18.03209-X

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA*, *295*(6), 676–680. https://doi.org/10.1001/jama.295.6.676

*Pizzighello, S., & Bressan, P. (2008). Auditory attention causes visual inattentional blindness. *Perception*, *37*(6), 859–866. https://doi.org/10.1068/p5723

Polanin, J. R., & Pigott, T. D. (2015). The use of meta-analytic statistical significance testing. *Research Synthesis Methods*, *6*(1), 63–73. https://doi.org/10.1002/jrsm.1124

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*(5), 2641–2651. https://doi.org/10.3758/s13414-020-01982-8

Pustejovsky, J. (2022). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections* (R Package Version 0.5.5) [Computer software]. https://CRAN.R-project.org/package=clubSandwich

Pustejovsky, J., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

*Redlich, D., Memmert, D., & Kreitz, C. (2022). Does hunger promote the detection of foods? The effect of value on inattentional blindness. *Psychological Research*, *86*(1), 98–109. https://doi.org/10.1007/s00426-021-01480-y

*Redlich, D., Schnuerch, R., Memmert, D., & Kreitz, C. (2019). Dollars do not determine detection: Monetary value associated with unexpected objects does not affect the likelihood of inattentional blindness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *72*(9), 2141–2154. https://doi.org/10.1177/1747021819835148

Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999). Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science*, *286*(5449), 2504–2507. https://doi.org/10.1126/science.286.5449.2504

*Remington, A., Cartwright-Finch, U., & Lavie, N. (2014). I can see clearly now: The effects of age and perceptual load on inattentional blindness. *Frontiers in Human Neuroscience*, *8*, Article 229. https://doi.org/10.3389/fnhum.2014.00229

Richards, A., Hannon, E. M., & Derakshan, N. (2010). Predicting and manipulating the incidence of inattentional blindness. *Psychological Research*, *74*(6), 513–523. https://doi.org/10.1007/s00426-009-0273-8

*Rivardo, M., Brown, K., Rodgers, A., Maurer, S., Camaione, T., Minjock, R., & Gowen, G. (2011). Integrating inattentional blindness and eye witness memory. *North American Journal of Psychology*, *13*, 519–538.

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, *26*(2), 141–160. https://doi.org/10.1037/met0000300

Rothstein, H., Sutton, A., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. https://onlinelibrary.wiley.com/doi/book/10.1002/0470870168

Schnuerch, R., Kreitz, C., Gibbons, H., & Memmert, D. (2016). Not quite so blind: Semantic processing despite inattentional blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 459–463. https://doi.org/10.1037/xhp0000205

Simons, D. J. (2010). Monkeying around with the gorillas in our midst: Familiarity with an inattentional-blindness task does not improve the detection of unexpected events. *i-Perception*, *1*(1), 3–6. https://doi.org/10.1068/i0386

*Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. https://doi.org/10.1068/p281059

*Simons, D. J., & Jensen, M. S. (2009). The effects of individual differences and task difficulty on inattentional blindness. *Psychonomic Bulletin & Review*, *16*(2), 398–403. https://doi.org/10.3758/PBR.16.2.398

Simons, D. J., & Schlosser, M. D. (2017). Inattentional blindness for a gun during a simulated police vehicle stop. *Cognitive Research: Principles and Implications*, *2*(1), Article 37. https://doi.org/10.1186/s41235-017-0074-3

Srinivasan, N., & Gupta, R. (2010). Emotion-attention interactions in recognition memory for distractor faces. *Emotion*, *10*(2), 207–215. https://doi.org/10.1037/a0018487

*Stothart, C., Simons, D. J., Boot, W. R., & Wright, T. J. (2019). What to where: The right attention set for the wrong location. *Perception*, *48*(7), 602–615. https://doi.org/10.1177/0301006619854302

*Swettenham, J., Remington, A., Murphy, P., Feuerstein, M., Grim, K., & Lavie, N. (2014). Seeing the unseen: Autism involves reduced susceptibility to inattentional blindness. *Neuropsychology*, *28*(4), 563–570. https://doi.org/10.1037/neu0000042

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

*Thomas, F. M. F., & Charlton, S. G. (2020). Inattentional blindness and information relevance of variable message signs. *Accident Analysis and Prevention*, *140*, Article 105511. https://doi.org/10.1016/j.aap.2020.105511

*Todd, J. J., Fougnie, D., & Marois, R. (2005). Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentional blindness. *Psychological Science*, *16*(12), 965–972. https://doi.org/10.1111/j.1467-9280.2005.01645.x

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2019). Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLOS ONE*, *14*(4), Article e0215052. https://doi.org/10.1371/journal.pone.0215052

van Boxtel, J. J., Tsuchiya, N., & Koch, C. (2010). Opposing effects of attention and consciousness on afterimages. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8883–8888. https://doi.org/10.1073/pnas.0913292107

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

*Velichkovksy, B., & Popova, S. (2021). Gender differences in object and spatial inattentional blindness under working memory load. In B. M. Velichkovsky, P. M. Balaban, & V. L. Ushakov (Eds.), *Advances in cognitive research, artificial intelligence and neuroinformatics* (Vol. 1358, pp. 122–127). Springer. https://doi.org/10.1007/978-3-030-71637-0_14

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

*Wei, X., Zhang, H., Hu, J., Xu, J., & Wang, J. (2021). The effect of emotional valence and age of faces on adults and children's inattentional blindness. *Attention, Perception & Psychophysics*, *83*(4), 1571–1580. https://doi.org/10.3758/s13414-020-02221-w

*White, R. C., & Davies, A. A. (2008). Attention set for number: Expectation and perceptual load in inattentional blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1092–1107. https://doi.org/10.1037/0096-1523.34.5.1092

White, R. C., Davies, M., & Aimola Davies, A. M. (2018). Inattentional blindness on the full-attention trial: Are we throwing out the baby with the bathwater? *Consciousness and Cognition*, *59*, 64–77. https://doi.org/10.1016/j.concog.2017.10.002

*Wiemer, J., Gerdes, A. B., & Pauli, P. (2013). The effects of an unexpected spider stimulus on skin conductance responses and eye movements: An inattentional blindness study. *Psychological Research*, *77*(2), 155–166. https://doi.org/10.1007/s00426-011-0407-7

Wolfe, J., & Horowitz, T. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*, Article 0058. https://doi.org/10.1038/s41562-017-0058

*Wood, K., & Simons, D. J. (2017a). Selective attention in inattentional blindness: Selection is specific but suppression is not. *Collabra: Psychology*, *3*(1), Article 19. https://doi.org/10.1525/collabra.90

*Wood, K., & Simons, D. J. (2017b). The role of similarity in inattentional blindness: Selective enhancement, selective suppression, or both? *Visual Cognition*, *25*(9–10), 972–980. https://doi.org/10.1080/13506285.2017.1365791

*Wood, K., & Simons, D. J. (2019). *The role of attention set in the timing of noticing in inattentional blindness* [Unpublished raw data]. https://osf.io/c3axj/

Wright, T. J., Boot, W. R., & Morgan, C. S. (2013). Pupillary response predicts multiple object tracking load, error rate, and conscientiousness, but not inattentional blindness. *Acta Psychologica*, *144*(1), 6–11. https://doi.org/10.1016/j.actpsy.2013.04.018

*Yan, X., Liu, B., Zhang, X.-M., Wei, L.-Q., & Zhao, X. (2012). Mechanism of distractor processing in dynamic inattentional blindness: Is there distractor inhibition? *Acta Psychologica Sinica*, *44*(5), 595–604. https://doi.org/10.3724/SP.J.1041.2012.00595

*Zhang, H., Wang, J., Liu, Y., Yan, C., & Ye, X. (2022). Threat-relevant stimuli cannot be better detected by preschoolers in an inattentional blindness task. *Psychological Research*, *86*(3), 823–830. https://doi.org/10.1007/s00426-021-01530-5

*Zhang, H., Yan, C., Zhang, X., & Fang, J. (2018). Sustained inattentional blindness does not always decrease with age. *Frontiers in Psychology*, *9*, Article 1390. https://doi.org/10.3389/fpsyg.2018.01390

Received September 8, 2021
Revision received July 12, 2022
Accepted July 18, 2022 ■