Inference about Absence as a Window into the Mental Self-Model

Matan Mazor[1]

[1] Wellcome Centre for Human Neuroimaging

Author Note

Correspondence concerning this article should be addressed to Matan Mazor, . E-mail: mtnmzor@gmail.com

Abstract

To represent something as absent, one must know that they would have known if it was present. This form of counterfactual reasoning critically relies on a *mental self-model*: a simplified schema of one's own cognition, which specifies expected perceptual and cognitive states under different world states and affords better monitoring and control over cognitive resources. Here I propose to use inference about absence as a unique window into the structure and function of the mental self-model. In contrast to commonly used paradigms, using inference about absence bypasses the need for explicit metacognitive reports. I draw on findings from low-level perception, spatial attention, and episodic memory, in support of the idea that self knowledge is a computational bottleneck for efficient inference about absence, making inference about absence a cross-cutting framework for probing key features of the mental self-model that are not accessible for introspection.

Keywords: self-model, absence, metacognition

Word count: 5800

Inference about Absence as a Window into the Mental Self-Model

You are in the grocery shop. On your grocery list are one carton of oat milk and one guava. You search through the shelves and find your favourite oat milk. You place the carton in your basket and move on to the fruit aisle. You visually scan the fruit boxes, but you already have a strong feeling that you will not find guavas in this store. You would have already smelled the guavas if they were anywhere around you. But then again, maybe something is wrong with your sense of smell? You grab a mandarin and sniff it. Your sense of smell is intact. You can be confident that there are no guavas around.

Inference about absence

Finding the oat milk carton was straightforward. As soon as you identified it you were convinced in its presence, no reflection or deliberation required. In contrast, concluding that no guavas were present took you longer and involved more complex cognitive processes. You had to rely on the absence of smell or sight of the fruit to reach a conclusion. In philosophical writings, this is known as Argument from ignorance (*Argumentum ad ignorantiam*): the fallacy of accepting a statement as true only because it hasn't been disproved (Locke, 1836). Although logically unsound, *Argumentum ad ignorantiam* is widely applied by humans in different situations and contexts (Oaksford & Hahn, 2004). One particular context which invites such reasoning is that of inference about absence. Positive evidence is rarely available to support inference about absence, and so it is often made on the basis of a failure to find evidence for presence.

Basing inference on the absence of evidence can sometimes be rational from a Bayesian standpoint (Oaksford & Hahn, 2004). For this to be the case, the individual must know the sensitivity and specificity of the perceptual or cognitive system at hand. For example, in order for the inference "I don't smell a guava, therefore there are no guavas in this store" to be logically sound, I need to know that the probability of me not smelling a guava is very low if it is nearby, and so is the probability of me imagining the smell of a guava when it is not there. In other words, in order to make valid inferences about absences I need to know things about myself and my cognitive processes. In the above example, this is evident in that my certainty in the absence of a guava increased after smelling the mandarin. Critically, smelling the mandarin did not provide me with any additional information about the layout of the shop or the seasonal availability of tropical fruit, but about my own perceptual system.

This example of inference about absence is exceptional in that I am able to justify my reasoning. If later my friend asks me why I concluded that no guavas were in the store, I will be able to convince them by explaining how I normally smell guavas from a distance, how I was able to smell the mandarin, and how I concluded that I would have detected a guava if it was present. But explicitly representing a derivation chain from assumptions to conclusions is the exception, not the rule. I can tell with confidence that there is no cup of water on my desk right now. If my friend asks me how I concluded that there was no cup of water on my desk, I would probably answer that I could see that it was not there. But this does not mean that I perceived its absence. It means that I did not perceive its presence, and that I would see if it was there. The first part is a fact about my perception, but the second part is based on intricate knowledge that details how hypothetical glasses of water may look like to me if they were on my desk right now. This builds on my knowledge of glasses, but more relevant to us here, on a *mental self-model*: a simplified description of one's own cognition, perception and attention that allows agents to predict their mental states under different world states.

Here I argue that this necessary role of a mental self-model for inference about absence makes inference about absence a promising tool to probe people's self-knowledge. Beliefs about my sense of smell, or the expected appearance of cups of water, are only part of a rich and complex knowledge structure, comprising beliefs about the senses (for example, the belief that my hearing is better in the right ear), attention (that I'm easily distracted by noises), and cognition (that I have bad memory for faces). Indeed, mental self-models have been suggested to play an important role in attention control (Wilterson et al., 2020), theory of mind (Graziano, 2019), and subjectivity more generally (Metzinger, 2003). While I can report some of those beliefs, some are not available to report, potentially not even to introspection (Flavell, 1979). This cognitive impenetrability is not different from how grammar rules are represented in cognition. Native English speakers would agree that the question “Who did you see Mary with?” is grammatically acceptable and that the question “Who did you see Mary and?” is not (Ross, 1967), but most would not be able to tell what rule is violated by the second question. Similarly, one may immediately appreciate that an object is missing, even if they will not be able to provide a better justification for this impression other than “I could see that it was not there”.

The following section introduces a computational formulation of this self-knowledge account, based in formal semantics and Bayesian theories of cognition, and exemplifies how different patterns of results can be interpreted in light of this formulation. This formulation is then followed by descriptions of several independent lines of experimental work that all share a role for self-knowledge in inference about absence. Finally, I present a vision for how future work can utilize these mechanisms to learn about the structure of this knowledge and about its acquisition over the course of development.

Probabilistic reasoning, criterion setting, and self knowledge

This paper is not the first to point out the intimate link between inference about absence and self-knowledge. In *default-reasoning logic* (Reiter, 1980), a failure to provide a proof for a statement is transformed into a proof for the negation of the statement using the *closed world assumption*: the assumption that a proof would have been found if it was avaiable. Similarly, Linguist Benoît de Cornulier’s refers to *epistemic closure*: the notion that all there is to be known is in fact known. This is reflected in his two definitions of *knowing whether* (De Cornulier, 1988):

Symmetrical definition:

"John knows whether P" means that:

1. If P, John knows that P.
2. If not-P, John knows that not-P.

Dissymmetrical definition:

"John knows whether P" means that:

1. If P, John knows that P.
2. John knows that 1 holds.

The symmetrical definition is available for statements that can be supported or negated by evidence. For example, the statement "It is not yet 3pm" can be supported if the time on one's phone indicates that it is 2:30pm, or negated if the time on one's phone indicates it is 3:30pm. Therefore, knowing whether it is now 3pm does not rely on self-knowledge. Conversely, statements such as "I have met this person before" can only be supported by positive evidence. This leaves inference about their negation to be made based on the absence of evidence, in conjunction with self-knowledge ("I don't recall seeing this person before, and this is not a face that I would forget"). This is an example of the De Cornulier's dissymmetrical definition: knowing that I would not have forgotten this person's face is in this case "knowing that 1 holds".

In psychological experiments of near-threshold detection, participants are required to decide whether a stimulus (for example a faint dot) was present or absent from a display. Using De Cornulier's formulation, we can ask which of the two definitions better describes the inferential machinery that is engaged in such tasks. Is it the case that participants perceive positive evidence for he absence of a target (symmetrical definition), or alternatively, do they rely on the metacognitive belief that they would have seen the target if it was present (dissymetrical definition)?

The *high-threshold model* of visual detection (Blackwell, 1952) formalizes this process in a way that shares conceptual similarity with De Cornulier's dissymemetrical definition. According to the high-threshold model, the probability of detecting the signal *d* scales with stimulus intensity. If participants detect the signal, they respond with "yes". The parameter *d* is a perceptual parameter: it captures variables such as objective stimulus intensity (for example, in units of luminance) and sensory sensitivity (for example, of photoreceptors in the retina, or neurons in the visual cortex). The value of this parameter corresponds to the degree to which statement 1 in the dissymetrical definition is true: "If P [a stimulus is presented] John knows that P". Critically, in the high-threshold model no similar parameter exists to control the probability of detecting the absence of a signal. In other words, the presence/absence asymmetry is expressed in the absence of a direct edge from "stimulus absent" to a "no" response (leftmost dashed line in Fig. 1, upper panel). In this model, "no" responses are controlled by the "guessing" parameter *g*. Unlike *d*, the *g* parameter is under participants' cognitive control, and can be optimally set to maximize accuracy based on beliefs about the probability of a stimulus to appear, the incentive structure, and critically, metacognitive beliefs about the perceptual sensitivity parameter *d*.

The high-threshold model, like other discrete state accounts of perception, has mostly been neglected in light of evidence of graded perception, even for sub-threshold stimuli (e.g., Koenig & Hofer, 2011). Still, continuous and graded models of perception based on Signal Detection Theory (SDT) express the same asymmetrical nature of presence/absence judgments, where clear evidence can be available for presence but less so for absence. In signal detection terms, this is expressed as high between-trial variance in sensory strength when a signal is present, but low variance when a signal is absent (see Fig. 1, lower panel). Here, instead of controlling the parameter *g*, participants control the placement of a decision criterion. Only trials in which the sensory signal (also termed perceptual evidence, or

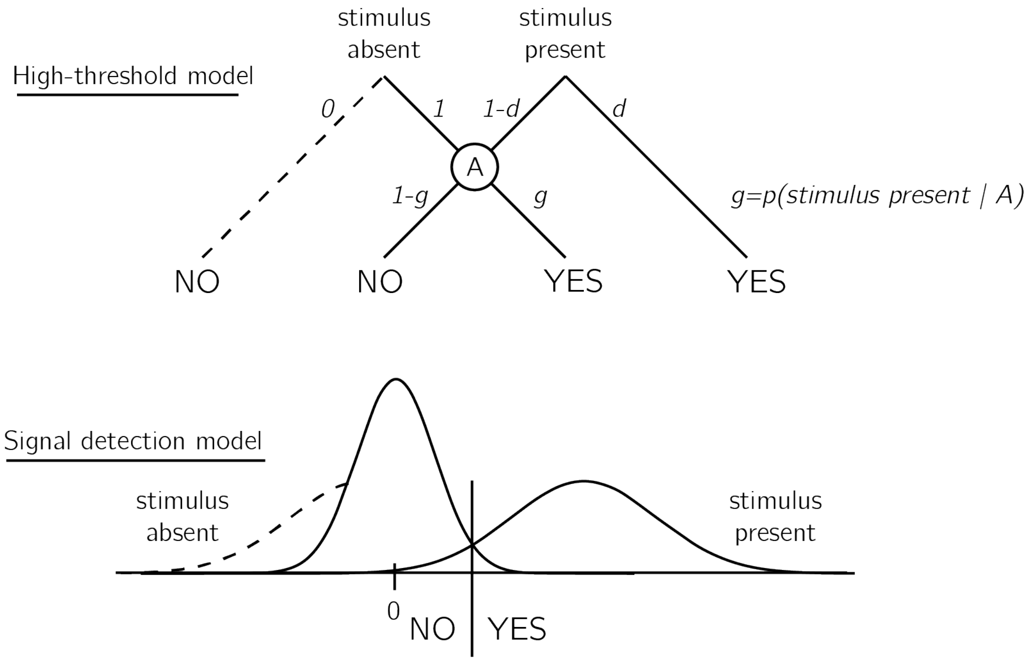


Figure 1. State and strength models of detection, commonly used in visual perception and recognition memory. In discrete high-threshold models (upper panel), the presence of a signal can sometimes lead directly to a 'yes' response, but the absence of a signal is never sufficient to lead to a 'no' response. 'No' responses are controlled by the parameter *g* - a 'guessing parameter' that determines the probability of responding 'yes' in case no stimulus was detected. In unequal-variance SDT models (lower panel), decisions are made based on the relative position of the sensory sample to a decision criterion. The presense/absence asymmetry manifests in the fact that only in some 'target-present' trials, but no in 'target-absent' trials, the sensory sample falls far away from the decision criterion. In both cases, given accurate prior knowledge about the prior probability of signal presence and the incentive strucutre, the *g* parameter and the decision criterion can be optimally set based on beliefs about one's own sensitivity to signal.

decision variable) exceeds this criterion will be classified as “stimulus present” trials. Optimal positioning of the criterion is dependent on beliefs about the likelihood of a stimulus to be present, as well as the spread of the signal and noise distributions and the distance between them. Due to the unequal-variance structure, sensory strength in trials where a stimulus is present will be on average farther from the decision criterion compared to when no stimulus is present. As a result, similar to the setting of the g parameter in the high-threshold model, the exact placement of the SDT decision criterion will have a bigger effect on accuracy when a stimulus is absent, compared to when a stimulus is present. In other words, the metacognitive process of setting the SDT criterion is critical for inference about absence much more than for inference about presence.

Common to both frameworks is the reliance on knowledge about one’s own perception (the d parameter in the first case, the shape and position of the sensory distributions in the second) for optimally setting a heuristic for response on trials in which no clear evidence is available for the presence of a signal. As a result, these models draw a strong link between participants’ beliefs about their own perception and their behaviour on target-absent trials. In what follows I provide empirical examples for how humans make inference about the absence of objects and memories. I link those examples to the core idea, that inference about absence critically relies on access to a self-model. Finally, I demonstrate how this link can be utilized by researchers to investigate participants mental (perceptual and cognitive) self-models.

Detection: “I would have noticed it”

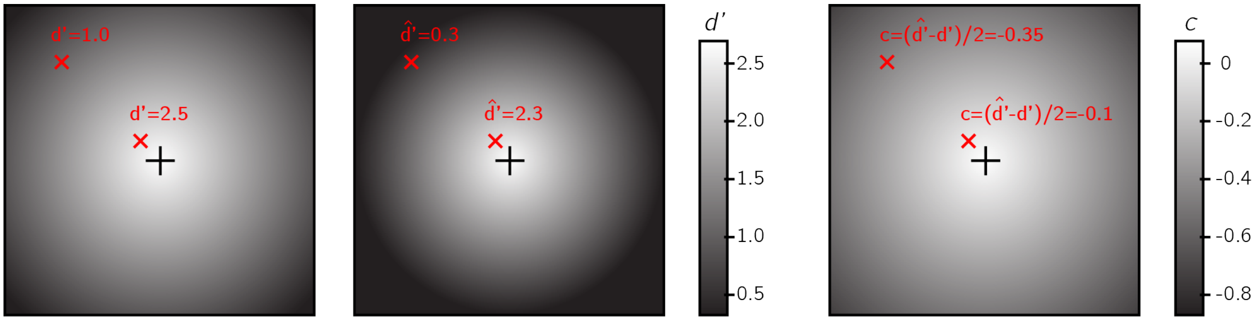
We start our exploration of inference about absence in cognition with perhaps the most basic of psychophysical tasks, visual detection. In visual detection, participants report the presence or absence of a target stimulus, commonly presented near perceptual threshold. In such tasks, accuracy alone cannot reveal a difference in processing between decisions about presence and decisions about absence, because task accuracy is a function of both “yes” and "no" responses.

However, when asked to report how confident they are in their decision, subjective confidence reports reveal a metacognitive asymmetry between judgments about presence and absence. Decisions about target absence are accompanied by lower confidence, even for correctly rejected "stimulus absence" trials (Kanai, Walsh, & Tseng, 2010; Mazor, Friston, & Fleming, 2020; Meuwese, Loon, Lamme, & Fahrenfort, 2014). Put differently, often participants cannot tell if they missed an existing target, or correctly perceived the absence of a target. A similar pattern is observed for response times: decisions about absence tend to be slower than decisions about presence (Mazor et al., 2020).

These observations fit well with the high-threshold and unequal-variance SDT models described above. Only in the presence of a target stimulus can participants make a decision without deliberation (without passing in the *A* node in the high-threshold model, or based on a sample very far from the decision criterion in unequal-variance SDT). On these trials, participants can be highly confident in that a target was present – more confident than when deciding that a target was present after deliberation. These high-confidence trials will not be available for decisions about target-absence.

In line with a central role for self-monitoring in inference about absence, this metacognitive blindspot for "stimulus absence" judgments diminishes or reverses when targets are masked from awareness by means of an attentional manipulation (Kanai et al., 2010; Kellij, Fahrenfort, Lau, Peters, & Odegaard, 2018). For example, when an attentional-blink paradigm is used to control stimulus visibility, participants are significantly more confident in their correct rejection trials than in their misses. What is it in attentional manipulations that improves participants' metacognitive insight into their judgments about stimulus absence? One compelling possibility is that a blockage of sensory information at the perceptual stage is not accessible to awareness (and is thus phenomenally transparent; Metzinger, 2003), whereas fluctuations in attention are accessible to introspection (and are thus phenomenally opaque; Limanowski & Friston, 2018). This monitoring of one's attention state makes it possible to use premises such as "I would not have missed the target" in rating confidence in absence under attentional, but not under perceptual manipulations of visibility. Put in more formal terms, attentional manipulations increase metacognitive access to the likelihood function going from world-states to perceptual states, thereby allowing trial-to-trial tuning of the decision criterion or the g parameter.

Studies contrasting detection responses and confidence ratings under different levels of attention provide more support for this metacognitive account of detection "no" responses. For example, participants are more likely to report the absence of a target in a specific location if their attention was directed to this location before stimulus onset, compared to when their attention was directed to a different location (Rahnev et al., 2011). Similarly, participants are more likely to correctly report the absence of a target embedded in a stimulus (for example, a grating embedded in noise) when the stimulus is presented at the center of their visual field, compared to the periphery (Odegaard, Chang, Lau, & Cheung, 2018; Solovey, Graney, & Lau, 2015). Note that both effects are the exact opposite of what is expected based on that attention boosts sensory gain (Parr & Friston, 2019), because an increase in sensory gain without a change to the decision criterion would make false alarms, not correct rejections, more prevalent. They are however consistent with the idea that participants deploy a metacognitive strategy, shifting their decision criterion to accord with the expected strength of evidence given their current attentional state. If participants overestimate the effect of attention on their visual sensitivity, decision criterion, as measured in Signal Detection Theory, will be lower for attended versus unattended stimuli (see Fig. 2). Indeed, detection criterion is typically found to be lower for unattended stimuli (Odegaard et al., 2018; Rahnev et al., 2011; Solovey et al., 2015).



Peceptual sensitivity

Beliefs about perceptual sensitivity

Decision criterion

Figure 2. Left panel: Sensitivity to near-threshold stimuli is lower in the visual periphery. For example, d' equals 1.0 in top left of the screen, but is much higher near the center. Right panel: the perceptual decision criterion is lower (more 'yes' responses) in the visual periphery. Middle panel: if the effect of eccentricity on visual sensitivity is overestimated in participants' mental self-model (here d' in the top left corner is estimated to be 0.3), a lowering of the decision criterion in the visual periphery as observed in Odegaard et al. (2018) is expected.

Visual search: "I would have found it"

In visual search tasks, participants are presented with an array of stimuli and are asked to report, as quickly and accurately as possible, whether a target stimulus was present or absent in the array. Moving one step up the complexity ladder, the accumulation of information in visual search is not only a function of stimulus strength and sensory precision, but is also affected by the endogenous allocation of attention to items in the visual array. As a result, search time varies as a function of the number of distractors, their perceptual similarity to the target and their spatial arrangement, among other factors (for a review, see Wolfe & Horowitz, 2008). These factors affect not only the time taken to report the presence of a target, but also the time taken to report its absence. For example, when searching for an orange target among red and green distractors, the number of distractors has virtually no effect on search time (e.g., D'Zmura, 1991) - a phenomenon known as "pop-out". The bottom-up pop-out of a target can explain the immediate recognition of the presence of a target, irrespective of distractor set size. But this perceptual pop-out cannot, by itself, explain the immediate recognition of target absence, because in target absence trials there is nothing in the display to pop out.

Computational models of visual search provide different accounts for search termination in target-absent trials. For example, in some versions of the *Guided Search* model, "target absent" judgments were the result of exhausting the search on items that surpassed a learned "activation threshold" (Chun & Wolfe, 1996; Wolfe, 1994). In difficult searches, the activation threshold was set to a low value, thereby requiring the scanning of multiple items before a "no" response can be delivered. In contrast, in easy searches the activation threshold could be set to a high value, reflecting a belief that a target would be highly salient. More recent models included a *quitting unit* that can be chosen with a certain probability (Moran, Zehetleitner, Müller, & Usher, 2013) or a *quitting threshold* parameter that resembles a noisy timer on search duration (Wolfe, 2021). Importantly for our point here, these different parameters all share high similarity with the SDT criterion or the high-threshold g parameter, and reflects explicit or implicit beliefs about the subjective salience of a hypothetical target in the array – a form of self-knowledge.

Usually, search times in target-present and target-absent trials are highly correlated, such that if participants take longer to find the target in a given display, they will also take longer to conclude that it is absent from it (Wolfe, 1998). This alignment speaks to the accuracy of the mental self-model: participants take longer to conclude that a target is missing when they belief they would take longer to find the target, and these beliefs about hypothetical search times are generally accurate. In the two upper panels of Fig. 3 I provide two examples of cases where beliefs about search behaviour perfectly align with actual serach behaviour, leading to optimal search termination. However, self-knowledge about attention in visual search is not always accurate. For example, when searching for an unfamiliar letter (for example, an inverted N) among familiar letters (for example, Ns), the unfamiliar letter draws immediate attention without a need for serially attending to each item in the display. However, participants are slow in concluding that no unfamiliar letter is present, exhibiting a search time pattern consistent with a serial search for “target absent” responses only (Wang, Cavanagh, & Green, 1994; Zhang & Onyper, 2020). In the context of my proposal here, this can be an indication for a blind-spot of the mental self-model, failing to represent the fact that an unfamiliar letter would stand out (see Fig. 3, lower panel).

Importantly, collecting explicit metacognitive judgments of expected search times may lead to underestimating the richness and accuracy of the mental self-model. For example, participants may not have introspective access to their knowledge about color pop-out, but may still be able to act on this information when deciding to terminating their search. Here also, inference about absence provides a unique window into the mental self-model that does not depend on introspective access.

Memory: “I would have remembered it”

Inference about absence not only applies to external objects (such as guavas, or visual items on the screen), but also to mental variables such as memories and thoughts. For example, upon being introduced to a new colleague, one can be certain that they have not met this person before. In the memory literature, this is known as *Negative recognition*: remembering that something did not happen (Brown, Lewis, & Monk, 1977). In the lab, a typical recognition memory experiment comprises a learning phase and a test phase. In the learning phase participants are presented with a list of items, and in the test phase they are asked to classify different items as “old” (presented in the learning phase) or “new” (not presented in the learning phase).

The role of self-knowledge in negative recognition is exemplified in the *mirror effect*: items that are more likely to be correctly endorsed as “old” are also more likely to be correctly rejected as “new”. For example, Brown et al. (1977) found that when asked to memorize a list of names, subjects are more confident in remembering that their own name was on the list, but also in correctly remembering when it was *not* on the list. For this effect

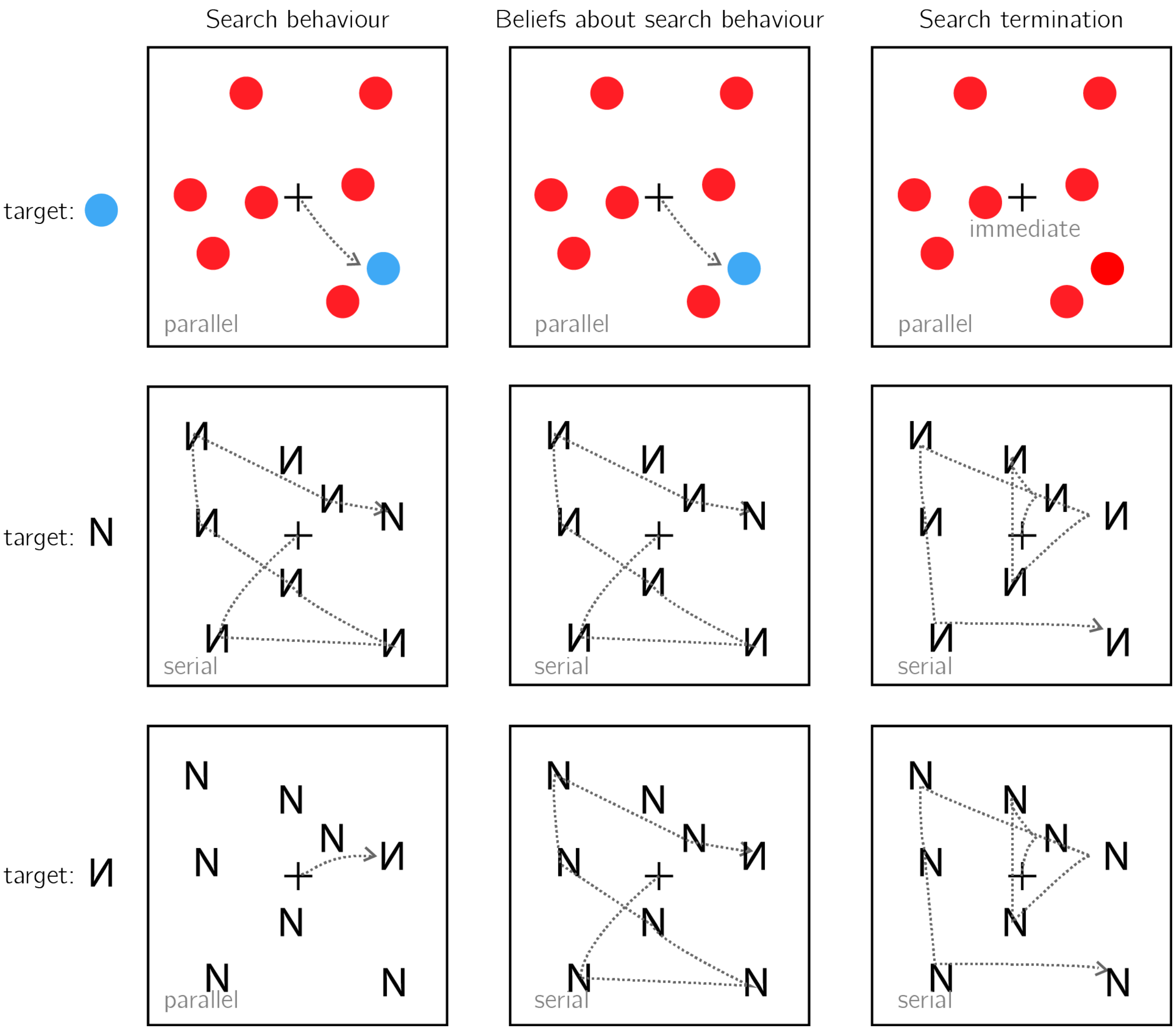


Figure 3. Upper panel: A target that is marked by a unique colour imemdiately captures attention (left). This fact is available to particiapnts' self-model (middle). As a result, participants can immediately terminate a search when no distractor shares the color of the target (right). Middle panel: When searching for the letter N among inverted Ns, the target does not immediately capture attention, and the serial deployment of attention is necessary (left). Participants are aware of this (middle). As a result, participants perform an exhaustive serial search before concluding that a target is absent (right. Lower panel: When searching for an inverted N among canincally presented Ns, the inverted letter immediately captures attention (left). This fact is not specified in the self-model (middle). As a result, participants perform an unncessary exhaustive serial search before concluding that a target is absent (right).

to manifest, it is not sufficient that subjects' memory was better for their own name. They also had to know this fact, and to use it in their counterfactual thinking ("I would have remembered if my name was on the list"). The mirror effect has also been demonstrated for the name of one's hometown (Brown et al., 1977), for word frequency (rare words are more likely to be correctly endorsed or rejected with confidence; Brown et al., 1977; Glanzer & Bowles, 1976), word imaginability (Cortese, Khanna, & Hacker, 2010; Cortese, McCarty, & Schock, 2015) and for study time (subjects are more likely to correctly reject items if learned items are presented for longer; Stretch & Wixted, 1998; Starns, White, & Ratcliff, 2012).

In a clever set of experiments, Strack, Förster, and Werth (2005) established a causal link from metacognitive beliefs about item memorability and decisions about the absence of memories. In two experiments, participants in one group were led to believe that high-frequency words (words that are used relatively often) are more memorable than low-frequency words, while participants in a second group were led to believe that low-frequency words were more memorable than high-frequency words. This manipulation affected participants' tendency to reject high-frequency or low-frequency items in a later recognition-memory task. Participants who believed that high-frequency words were more memorable were more likely to classify high-frequency words as "new", suggesting that their metacognitive belief informed their inference about the absence of a memory ("I would have remembered this word"). Inversely, participants who believed that low frequency words were more memorable showed the opposite pattern.

Just like in the cases of near-threshold detection and visual search, the intuitive metacognitive knowledge behind the mirror effect may not be available for explicit report, at least not in the absence of direct experience with the task itself. In their explicit memorability reports, subjects often have little to no declarative metacognitive knowledge of which items are more likely to be remembered, even under conditions that give rise to a mirror effect. For example, although more frequent words are more likely to be forgotten (and incorrectly classified as old), participants tended to judge them as more memorable than infrequent words (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin, 2003; Greene & Thapar, 1994; Wixted, 1992). However, participants showed metacognitive insight into the negative effect of word frequency on memorability when memorability was rated after (and not before) negative recognition judgments (Benjamin, 2003; Guttentag & Carroll, 1998). Thus, the implicit metacognitive knowledge that supports accurate negative recognition may become available for explicit report only when participants introspect over their recognition attempts. Similar to beliefs about perceptual sensitivity in the visual periphery or beliefs about attention in visual search, this may be one example where using inference about absence to probe self-knowledge can reveal more than what can be measured with explicit subjective reports.

The development of a self-model

As exemplified above, the inferential processes that result in judgments of absence share important commonalities, regardless of whether it is the absence of an isolated target stimulus, of one target in an array of distractors, or of a non-physical entity such as a memory. First, in all three cases, to infer absence agents must possess some self-knowledge (under what conditions are they likely to miss a target, how long they should expect to search before finding a target in an array of distractors, or which items are likely or unlikely to be remembered). Second, agents must be able to use this counterfactual knowledge and compare it with their current state (for example, having no recollection of an item, or not seeing a target stimulus).

At what developmental stage do humans master the necessary self knowledge and inferential machinery to make efficient and accurate inference about absence? In the context of memory, evidence suggests that the necessary self-knowledge and the capacity for counterfactual thinking exist in primary form already in early childhood, but continue to develop until adulthood. For example, children as young as 5 were able to give meaningful assessments the memorability of hypothetical life events and to use this metacognitive knowledge to inform their judgments about the nonoccurrence of an event, but this ability did not reach full maturation until the age of 9 (Ghetti & Alexander, 2004). Other studies identified a qualitative transition between the ages 7 and 8 in the ability of children to rely on expected event memorability for inference about the absence of a memory (Ghetti, Castelli, & Lyons, 2010; Ghetti, Lyons, Lazzarin, & Cornoldi, 2008). This developmental discontinuity was attributed to the development of counterfactual thinking and second-order theory of mind. Indeed, the ability to infer that something did not happen based on that it would have been remembered critically relies on one's ability to ascribe mental states to their counterfactual self.

In perception, the ability to represent absences lags behind the ability to represent presences, but reaches maturation much earlier than in the case of memory. In a study by Coldren and Haaf (2000), 4 month-old infants were familiarized with a pair of identical letters (e.g., the letter "O"), presented side by side. In the test phase, one of the letters was replaced with a novel letter, which differed from the familiar letter either in the presence or the absence of a distinctive feature. For example, when infants that were familiarized with the letter O were tested on a display of one O and one Q, the novel letter (Q) was marked by the presence of a distinctive feature. Conversely, for infants that were familiarized with the letter Q, the novel letter O was marked by the absence of a distinctive feature. Infants showed preferential looking at the novel letter only when this letter was marked by the presence, not the absence, of a distinctive feature. A similar feature-positive effect was still evident in the learning behaviour of preschool children. When presented with two similar displays, 4 and 5 year old children were able to learn to approach the display with a distinctive feature but were at chance when trained to approach a display that is marked by the absence of a distinctive feature (Sainsbury, 1971).

Together, these results suggest that the capacity to infer the absence of physical and mental entities develops through infancy and early childhood. In context of the framework presented here, the development of this capacity can reflect the gradual expansion of different aspects a mental self-model, and the development of the capacity to use this model for counterfactual reasoning. For example, a baby that is not drawn to the new letter "O" after being habituated to the letter "Q" may not yet represent the absence of the distinguishing feature, because they lack the implicit self knowledge to know that they would have noticed the lower diagonal line if it was present. More abstractly, a 7 year-old may not be able to confidently tell that they did not spread a lotion on a chair (a highly memorable action, due to its bizarreness; Ghetti et al., 2008), because they lack the self-knowledge to know that if they had, they would have remembered doing so.

Using inference about absence to study the mental self-model

In this paper I argue that the mental self-model plays an important role in inference about absence. I provide examples from near-threshold perception, visual search, and recognition memory, for cases where accurate beliefs about one's own perception and cognition can increase the accuracy, speed, and metacognitive access to the quality of decisions about the absence of objects or memories. Examples from the developmental psychology literature indicate that these beliefs develop in infancy and childhood. In this last part I list practical directions for utilizing the reliance of inference about absence on self-knowledge to study the mental self-model.

Inverting the mental self-model

Our working assumption is that inference about absence draws on knowledge from a mental self-model. Given this assumption, behavioural markers of inference about absence (such as decision time, accuracy, and subjective confidence) can be used to answer the question "which specification of the mental self-model would give rise to this behaviour?". In other words, these measures can decide between competing mental self-models that subjects may have at the time of performing the task. In the above examples, behaviour was used to identify qualitative properties of the self-model, such as an exaggerated effect of attention on perceptual sensitivity, or no knowledge of the immediate capturing of attention by unfamiliar stimuli. This approach can be taken one step further by specifying a model family and idenfying model parameters that agree with the observed data.

As an example, consider the effect of eccentricity on the decision criterion, described in the section about visual detection. In Fig. 2 I illustrate how biased beliefs about the effects on eccentricity on perceptual sensitivity (middle panel) can give rise to a shift in the decision criterion (right panel). For every pair $[d', \hat{d'}]$, this simple model optimally sets the decision criterion to maximize accuracy given the subjec'ts'"s estimate of their sensitivity $\hat{d'}$, and measures the resulting 'empirical criterion" with respect to their actual sensitivity d'. Importantly, given a pair of observed $[d', c]$ from a detection study, this model can then be inverted to obtain the $\hat{d'}$ that would give rise to the observed empirical criterion. Recently, Winter and Peters (2021) estimated qualitative metacognitive beliefs about perceptual noise levels from confidence ratings in a visual discrimination task. In order to obtain quantitative measures of these beliefs, future work will need to employ more systematic parameter recovery starting from empirical data (for example, using Markov-Chain Monte-Carlo sampling).

Quantifying implicit learning effects

Similar to models of the world or of one's body and motor system, a mental self-model is expected to expand and change in light of new evidence. I suggest that these changes to the self-model will be most evident in decisions about absence. For example, in discussing inference about absence in the context of memory, I described a study where participants were led to believe that high usage frequency made words more or less memorable (Strack et al., 2005). These beliefs were later reflected in participants' tendency to categorize high and low frequency words as "old" or "new". In one experiment, belief induction was obtained without explicitly telling participants which words were more memorable. Instead, Strack and colleagues made use of the fact that high-frequency words are more easily recalled in free-recall paradigms, but low-frequency words are more easily recognized in item recognition paradigms. An additional free-recall/item-recognition task prior to the main recognition memory test induced different beliefs about item memorability in the two experimental groups. These newly acquired beliefs reflected in participants' negative recognition judgments, without a need to explicitly probe participants' explicit metacognitive beliefs about word memorability.

Similarly, future studies can use a similar approach to induce beliefs about the relative difficulty of visual search tasks or the detectability of specific visual stimuli. For example, can participants be led to believe that searching for a T among Ls is easier than what is actually is (for example, by using subliminal spatial cuing immediately prior to the search array)? Would this belief extend to other letters, or geometrical shapes? How persistent would such a belief be (for example, would it still be evident in the target-absent search-times in the next day or week)?

Contrasting explicit and implicit self-knowledge

Metacognitive knowledge is typically probed in the lab by means of explicit report, for example, by asking subjects to rate their ability or make prospective confidence ratings (Fleming, Massoni, Gajdos, & Vergnaud, 2016). In this paper I provided examples for that some self-knowledge is accessible only to some subsystems, encapsulated from introspection. Extracting the contents of the mental self-model based on inference about absence may, in some cases, reveal self-knowledge that is not available for explicit report but is used to guide behaviour. This is similar to how grammaticality judgments reveal linguistic knowledge that is not available in the form of declarative knowledge. As exemplified above, this can be highly beneficial in the developmental study of babies and infants, who may not be able to provide reliable explicit metacognitive ratings due to limited communication skills or the lack of an explicit theory of mind, but whose implicit mental self-model is growing and changing in telling and interesting ways.

Conclusion

An accurate mental self-model can be useful in transforming the absence of evidence for the presence of objects or memories into beliefs about the absence of objects or memories. Findings from the fields of visual psychophysics and recognition memory suggest that this model is sometimes exaggerated or over-simplified, and that it develops with age and task experience. Here I suggest to utilize the tight link between inference about absence and the mental self-model to empirically study the structure and contents of this model, without assuming that participants have full access to it at all times.

Acknowledgements

I thank Stephen M. Fleming and Nadine Dijkstra for valuable feedback and comments, and Ezer Rasin and Maayan Keshev for their helpful input on nondeclarative linguistic knowledge.

References

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610–632.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, *31*(2), 297–305.

Blackwell, H. R. (1952). Studies of psychophysical methods for measuring visual thresholds. *JOSA*, *42*(9), 606–616.

Brown, J., Lewis, V., & Monk, A. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of Experimental Psychology*, *29*(3), 461–473.

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*(1), 39–78.

Coldren, J. T., & Haaf, R. A. (2000). Asymmetries in infants' attention to the presence or absence of features. *The Journal of Genetic Psychology*, *161*(4), 420–434.

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609.

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*, *68*(8), 1489–1501.

De Cornulier, B. (1988). Knowing whether, knowing who, and epistemic closure. *Questions and Questioning*, 182–192.

D'Zmura, M. (1991). Color in visual search. *Vision Research*, *31*(6), 951–966.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, *2016*(1).

Ghetti, S., & Alexander, K. W. (2004). "If it happened, i would remember it": Strategic use of event memorability in the rejection of false autobiographical events. *Child Development*, *75*(2), 542–561.

Ghetti, S., Castelli, P., & Lyons, K. E. (2010). Knowing about not remembering: Developmental dissociations in lack-of-memory monitoring. *Developmental Science*, *13*(4), 611–621.

Ghetti, S., Lyons, K. E., Lazzarin, F., & Cornoldi, C. (2008). The development of metamemory monitoring during retrieval: The case of memory strength and memory absence. *Journal of Experimental Child Psychology*, *99*(3), 157–181.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 21.

Graziano, M. S. (2019). Attributing awareness to others: The attention schema theory and its relationship to behavioural prediction. *Journal of Consciousness Studies*, *26*(3-4), 17–37.

Greene, R. L., & Thapar, A. (1994). Mirror effect in frequency discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 946.

Guttentag, R., & Carroll, D. (1998). Memorability judgments for high-and low-frequency words. *Memory & Cognition*, *26*(5), 951–958.

Kanai, R., Walsh, V., & Tseng, C.-h. (2010). Subjective discriminability of invisibility: A framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*, *19*(4), 1045–1057.

Kellij, S., Fahrenfort, J., Lau, H., Peters, M. A., & Odegaard, B. (2018). The foundations of introspective access: How the relative precision of target encoding influences metacognitive performance.

Koenig, D., & Hofer, H. (2011). The absolute threshold of cone vision. *Journal of Vision*, *11*(1), 21–21.

Limanowski, J., & Friston, K. (2018). “Seeing the dark”: Grounding phenomenal transparency and opacity in precision estimation for active inference. *Frontiers in Psychology*, *9*, 643.

Locke, J. (1836). *An essay concerning human understanding*. T. Tegg; Son.

Mazor, M., Friston, K. J., & Fleming, S. M. (2020). Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *Elife*, *9*, e53900.

Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences*, *2*(4), 353–393.

Meuwese, J. D., Loon, A. M. van, Lamme, V. A., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: Comparing metacognition for object detection and object categorization. *Attention, Perception, & Psychophysics*, *76*(4), 1057–1068.

Moran, R., Zehetleitner, M., Müller, H. J., & Usher, M. (2013). Competitive guided search: Meeting the challenge of benchmark rt distributions. *Journal of Vision*, *13*(8), 24–24.

Oaksford, M., & Hahn, U. (2004). A bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *58*(November 2015), 75–85. https://doi.org/10.1037/h0085798

Odegaard, B., Chang, M. Y., Lau, H., & Cheung, S.-H. (2018). Inflation versus filling-in: Why we feel we see more than we actually do in peripheral vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 20170345.

Parr, T., & Friston, K. J. (2019). Attention or salience? *Current Opinion in Psychology*, *29*, 1–5.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., De Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, *14*(12), 1513–1515.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, *13*(1-2), 81–132.

Ross, J. R. (1967). Constraints on variables in syntax.

Sainsbury, R. (1971). The "feature positive effect" and simultaneous discrimination learning. *Journal of Experimental Child Psychology*, *11*(3), 347–356.

Solovey, G., Graney, G. G., & Lau, H. (2015). A decisional account of subjective inflation of visual perception at the periphery. *Attention, Perception, & Psychophysics*, *77*(1), 258–271.

Starns, J. J., White, C. N., & Ratcliff, R. (2012). The strength-based mirror effect in subjective strength ratings: The evidence for differentiation can be produced without differentiation. *Memory & Cognition*, *40*(8), 1189–1199.

Strack, F., Förster, J., & Werth, L. (2005). "Know thyself!" The role of idiosyncratic self-knowledge in recognition memory. *Journal of Memory and Language*, *52*(4), 628–638.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1379.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics*, *56*(5), 495–500.

Wilterson, A. I., Kemper, C. M., Kim, N., Webb, T. W., Reblando, A. M., & Graziano, M. S. (2020). Attention control and the attention schema theory of consciousness. *Progress in Neurobiology*, *195*, 101844.

Winter, C. J., & Peters, M. A. (2021). Variance misperception under skewed empirical noise statistics explains overconfidence in the visual periphery. *bioRxiv*.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 681.

Wolfe, J. (2021). Guided search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*.

Wolfe, J., & Horowitz, T. S. (2008). Visual search. *Scholarpedia*, *3*(7), 3325.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*(1), 33–39.

Zhang, Y. R., & Onyper, S. (2020). Visual search asymmetry depends on target-distractor feature similarity: Is the asymmetry simply a result of distractor rejection speed? *Attention, Perception, & Psychophysics*, *82*(1), 80–97.