Synthese (2022) 200:25
https://doi.org/10.1007/s11229-022-03508-1

ORIGINAL RESEARCH

Recognizing why vision is inferential

J. Brendan Ritchie[1]

Received: 18 June 2021 / Accepted: 15 November 2021 / Published online: 22 February 2022

Abstract
A theoretical pillars of vision science in the information-processing tradition is that perception involves unconscious inference. The classic support for this claim is that, since retinal inputs underdetermine their distal causes, visual perception must be the conclusion of a process that starts with premises representing both the sensory input and previous knowledge about the visible world. Focus on this "argument from underdetermination" gives the impression that, if it fails, there is little reason to think that visual processing involves unconscious inference. Here an alternative means of support for this pillar is proposed, based on another foundational challenge for the visual system: recognizing invariant properties of objects in the environment even though anything we encounter is never seen exactly the same way twice. Explaining how the visual system solves this invariance problem requires positing visual processes that exhibit many commonalities with inductive inference. Thus, this novel "argument from invariance" reveals one way in which visual processing clearly involves unconscious inference.

Keywords Unconscious inference · Vision · Object recognition · Mental representation · Bayesian modeling

1 Introduction

A theoretical pillar of vision science in the information-processing tradition is that perception involves unconscious inference.[1] The classic support for this pillar is that, since retinal inputs underdetermine their distal causes, visual perception must be the conclusion of a process that starts with premises representing both the sensory input and previous knowledge about the visible world. Call this the *argument from under-*

[1] See: Aggelopoulos (2015), Barlow (1990), Epstein (1973), Fodor and Pylyshyn (1981), Gregory (1970), Hochberg (1981), Palmer (1999), Rock (1983). For its historical roots, see Hatfield (2002).

✉ J. Brendan Ritchie
j.brendan.w.ritchie@gmail.com

[1] Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, USA

determination. In more contemporary forms this argument has often been grounded in applications of Bayesian models in vision science and debate has then centered on how these models should be interpreted.[2] However, whatever form the argument takes, whether it goes through depends on showing that the visual processes that "solve" the underdetermination problem have qualities that are typically considered distinctive of cognition (Hatfield, 2002).

The fixation on underdetermination invites the impression that the argument and the pillar stand or fall together: if the argument tumbles, little remains to prop up the idea that visual processing involves unconscious inference. In what follows I offer an alternative, and more stable, base for the pillar. Besides underdetermination, another foundational challenge for the visual system is to track invariant features of the environment even though anything we encounter is never seen exactly the same way twice. This *invariance problem* is clearest in the case of object recognition, which requires representing objects as the same across transformations of viewpoint (DiCarlo et al., 2012). As I argue, this problem has many features that are diagnostic of inductive inference. In turn, standard explanations of object recognition posit unconscious processes in the visual system that overcome this problem. Therefore, the fact that unconscious processes are posited to explain how the visual system solves an induction problem shows that some aspects of visual processing involves unconscious inference. Call this the *argument from invariance*. In what follows I develop and defend this argument and conclude that it is better able to bear the weight of a key theoretical pillar of vision science.

The paper is structured as follows. In Sect. 2, I lay out my argumentative strategy for defending the pillar. In Sect. 3, I make a case for moving beyond the argument from underdetermination and its Bayesian variant. In Sect. 4, I present the argument from invariance using object recognition as a case study. In Sect. 5, I address some potential challenges. Section 6 concludes the paper.

2 Laying the groundwork for the argument

Plausibly a (theoretical) inference is a "reasoned change in view": we start with some beliefs, and through deliberation, end up revising what we believe, or perhaps how strongly we believe it (Boghossian, 2014; Harman, 1986; Kiefer, 2017). In Fig. 1A we cannot see what birds are nesting, but knowing why elevated nests are constructed on piles in waterways deduce that they are probably ospreys. However this pattern of deliberation manifests itself, it easily qualifies as a case of inference in the above sense. It is also at least partially inductive, since prior knowledge is being exploited. Contrast Fig. 1B where we immediately see the bird *as* an osprey. It is this second sort of case that is the focus of research on object recognition and that I wish to show also involves a process similar to inductive inference. But before doing so, clarity is needed on two fronts: first, on how I see the dialectic around unconscious inference; and second, on the argumentative strategy I intend to adopt.

[2] See: Clark (2013), Gładziejewski (2016), Hohwy (2013), Kiefer (2017), Orlandi (2016), Rescorla (2015, 2021).

Fig. 1 Two views of ospreys. **A** Two ospreys on an elevated nest. **B** An osprey in flight

2.1 Framing the dialectic

Often it is claimed that at issue is whether unconscious visual processing is *literally*, as opposed to metaphorically, inferential (Hatfield, 2002; Kiefer, 2017; Orlandi, 2014). There are two issues with this framing.

First, visual processing may involve unconscious inference in some senses but not in others. On the one hand, one might mean that seeing simply *is* thinking, though operating swiftly and outside of awareness. Historically, Ibn Al-Haytham (c. 965–1040)—and later Von Helmholtz (1867)—had this sense of unconscious inference in mind (Hatfield, 2002).[3] Contemporary vision scientists do not. Instead, they standardly posit unconscious information-processing that is proprietary to the visual system itself (e.g. Rock, 1983). On the other hand, given the historical connection between computation and deduction, the very idea of information-processing can perhaps be seen as a vindication of at least the spirit of earlier theorists. In which case, in the context of explaining visual processing, working vision scientists may treat "unconscious inference" as synonymous with "unconscious information-processing". One may insist the relevant notion is intermediary between these alternative, but it is not obvious that, once we move away from them, there is a single, privileged sense of unconscious inference that can be distilled as opposed to a multitude of plausible candidates.

Second, even if some aspect of unconsciously visual processing is literally inferential highlighting this fact may be explanatorily superfluous, or worse, misleading, as it accentuates similarities between seeing and thinking when it is the differences that may matter. This sentiment is well expressed by Kanizsa (1985, pp. 27–28):

> …the main problem of a theory of this kind, in my opinion, is that of not being able to suggest any advance, because it bears the risk of extinguishing the desire of investigating phenomena for which it has always ready a prefabricated explanation. From this point of view it is preferable to focus on the differences between seeing and thinking, because these, by indicating the possibility that the two classes of phenomena obey to different rules can set us on the road of discovering these rules.

[3] The idea of unconscious inference is one of many insights about vision first made by Ibn Al-Haytham (latinized *Alhazen*) that were later rediscovered (Howard, 1996).

In light of these two issues, my discussion will not focus on whether appeals to unconscious inference are literal, but on whether there are commonalities between seeing and thinking that are *explanatory* when it comes to particular visual phenomena (Kanizsa, 1985; Pylyshyn, 1999; Rock, 1983). One can treat these commonalities, collectively, as an explication, or operationalization, of a scientifically useful sense of unconscious inference, but the overlap with our commonsense intuitions about inference will only be partial.[4] Here is one straightforward way in which positing unconscious inference would be explanatory: the visual phenomenon *itself* has features that we consider diagnostic of some kind of inference. If that were that the case, then the commonalities between seeing and thinking reflected in an explanation would be a natural consequence of what is being explained. Such a strategy has two parts: first showing that the phenomenon has some of the diagnostic features of an induction problem; and second, showing that the unconscious processes that are posited to explain the phenomenon meet plausible requirements for an inferential solution to the problem. I elaborate on this strategy below.

2.2 Characterizing unconscious inductive inference

As typically understood, a "problem" for the visual system is a mapping from sensory input to perceptual output that is not yet understood. Somehow the visual system achieves this mapping and we would like to explain how. When might such a problem be similar in form to induction? I take the following to be a relatively uncontroversial description of a kind of inductive inference: a deliberative process that involves generalizing from past experience to form beliefs about a present circumstance. Figure 1A presents a case that is inductive in this way. This description also points to three diagnostic features that suffice for an operationalization of when an information-processing problem can be considered inductive. First, it is *diachronic* in that it concerns how we overtly represent the present in light of the past, and might change what we believe accordingly. Second, it is *empirical* in that it particularly concerns our representation of past experiences or acquired knowledge, which we bring to bear in drawing conclusions about present circumstances. Finally, it is *extrapolative* in the sense that when we generalize from past experience to novel circumstances there will typically be various ways in which the past and present circumstances differ. In what follows, I will call a phenomenon that has all these features an *induction problem* for the visual system.[5]

A "solution" to an information-processing problem is an explanation of how the visual system maps the target sensory inputs to the outputs. If the problem is an

[4] In this regard the present approach is similar to those present in discussions of whether quasi-technical notions of emotion (Griffiths, 1997) or innateness (Samuels, 2004) are explanatorily useful to cognitive science.

[5] All of these diagnostic features could be interpreted in a manner that does not require explicit representation. Instead, the information or knowledge from prior experiences is somehow "implicitly" represented in the operation of the visual system. However, this broader interpretation would not seem to describe a form of inferential process and is closer to the sort of metaphorical usages that have often been criticized (Hatfield, 2002; Orlandi, 2014). In the present discussion, I only consider these features in the more restricted sense that requires explicit mental representation of the inputs to the process.

inductive one, then there will be further requirements regarding the representations that are posited and the process that maps these representations to the resulting percept (cf. Hatfield, 2002).[6]

Regarding the representations, one must be of the environment derived from the sensory input, while others reflect prior knowledge gained from experience. These representations provide the equivalent of premises, the contents of which stand in evidential support to the concluding representation. Looking at Fig. 1A, one represents both the present state of affairs, but also retrieves other representations about the local fauna and the intended purpose of elevated nests in waterways. Together, these provide the evidence for the hypothesis that the birds are ospreys. These are furthermore *mental* representations and a reasonable expectation is that an unconscious solution to an induction problem will trade in them as well. Following previous discussions of unconscious inference (Mole & Zhao, 2016; Orlandi, 2014), I will assume that it suffices for a state of the visual system to be a mental representation if it has content that is: (i) *distal*, in the sense of being about properties of the external environment; and (ii) *robust* in the sense that the content stays the same even when it is tokened in the absence of what it represents (Fodor, 1990).[7]

Regarding the process, there are two requirements. The first is that the visual system transitions between the mental representations in a way that is plausibly inferential; That is, given a representation of the present sensory environment, there is a transition to a percept of the visible world in light of information afforded by representations related to past experiences. It is common to characterize inferential transitions as rule-following. While in paradigmatic cases of conscious deliberation this may require that an agent "takes" the premises to support the conclusion (Boghossian, 2014), others have suggested that, even in the case of cognition, deliberation can operate swiftly, automatically, and outside of conscious awareness (Quilty-Dunn & Mandelbaum, 2018; Wright, 2014). When unconscious in this way, it is simply a matter of our cognitive architecture that, given that the premises are represented, the conclusion is reliably represented as well. Following Quilty-Dunn and Mandelbaum (2018), I will call such operations "bare inferential transitions". The first requirement then is that a visual process involves some form of bare inferential transition from the

[6] There are two senses in which the diagnostic features I have enumerated might be thought to apply to the *solutions* of mapping problems, depending on how each is characterized. First, in Fig. 1A, we might want to explain how one comes to guess that the birds are ospreys, given the evidence available. The answer, or "solution", in this case, is that one has used inductive reasoning. Second, we might then seek to explain how this deliberation is achieved, from an information-processing perspective. In which case, the "problem" itself is a mapping achieved via inductive deliberation and the information-processing "solution" must also exhibit the features, assuming it explains (rather than explains away) this deliberation. It is this second sense of mapping problems/solutions, which I have in mind.

[7] The content must also presumably be *original*, in the sense of not being determined by convention or the intentions of a separate agent (Searle, 1983). Furthermore, the internal state of the visual system that is the vehicle for the content must serve a representational *function*, like being used by the visual system to stand-in for what it represents to aid in further information-processing or action (Ramsey, 2007). Here I take these conditions for granted and focus on the conditions of distality and robustness.

representation of the sensory input, along with prior knowledge, to a concluding percept.[8]

Second, the process must recruit some kind of long-term memory store, whereby information gained from prior experience is recorded, and can be retrieved for comparison with the present sensory input. Appealing to memory in this way is arguably latent within the very idea of inductive inference (Aggelopoulos, 2015; Fodor & Pylyshyn, 1981). For making an inference about the present from the past requires being able to represent the past in light of the present. In Fig. 1A, one cannot conjecture that the birds are ospreys without first retrieving from memory information about the nests and different birds that live in the area in order to generate the hypothesis.

To summarize, an information-processing phenomenon presents an induction problem for the visual system if it has the following diagnostic features: it is diachronic, empirical, and extrapolative. If an unconscious visual process that is posited to solve (i.e. explain) this problem also satisfies the above requirements on representation and process, then it follows that an inductive inference problem is solved by some aspect of unconscious visual processing that has many commonalities with inductive inference. In this sense, explaining the phenomenon will require positing a form of unconscious inference. Of course, this is not the only route by which one may show that some aspect of unconscious visual processing is inferential. A phenomenon may fail to satisfy the conditions I have laid out, yet be inferential in some other sense. For example, it may still qualify as a form of unconscious deductive or abductive inference. Though in such cases similar requirements on mental representation and inferential transitions would still apply. Similarly, the characterization of induction I have offered presumes a kind of learning process: that we acquire information about the world and extrapolate from that information to novel circumstances. Thus, it rules out the possibility of wholly innate forms of unconscious inductive inference, in so far as what is innate is not learned, though presumably, for any kind of information-processing, some aspect of it must be innate.[9] While these considerations highlight the limited scope of my strategy, they are a natural consequence of the fact that the underdetermination problem is typically characterized as involving inductive inference.

[8] Quilty-Dunn and Mandelbaum (2018, pp. 6–8) require that an inferential transition be not just rule-following but also "logic-obeying". The notion of logic-obeying they have in mind is tied to the idea of discursive representational formats in which a representation can be decomposed into a canonical constituent structure. Thus, they include the requirement that bare inferential transitions occur in virtue of the architecture of a system being sensitive to the constituent structure of the representations involved. I have excluded this requirement because the same notion of logic-obeying would seem to be inherent in the very idea of information-processing as a species of computation. For under a very general characterization, all computations operate in accordance with rules that are sensitive to only the constituent structure of the symbols over which they are defined (Piccinini & Scarantino, 2010). To put the point simply: if visual information-processing operations are carried out over mental representations they will have a discursive format.

[9] Matters may ultimately depend on the sense of "innateness" being employed or how one characterizes the debate between empiricist and nativist hypotheses, both of which are topics of discussion in their own right (Linquist, 2018). Here I assume that a psychological capacity is innate just in case it is not learned (Ritchie, 2020; Samuels, 2002) and that the debate concerns domain-specific vs domain-general learning processes in development (Margolis & Laurence, 2013). As to how much learning, or what style of learning, is required by my characterization of an induction problem, I remain agnostic. For example, it is compatible with the possibility of zero-shot learning constrained by inductive biases built into the visual system.

3 Undermining the argument from underdetermination

Given the groundwork laid down, how does the argument from underdetermination hold up? In this section I make the case that the underdetermination problem is a poor fit for the argumentative strategy described above. More specifically, it is not an induction problem for the visual system because it is not inherently diachronic. Therefore, if explaining how the visual system solves variants of the underdetermination problem involves positing a form of unconscious inductive inference, it is not because the sensory input is underdetermined by its distal cause. Furthermore, the same issue also arises for Bayesian variants of the argument.

3.1 Underdetermination is not (obviously) an induction problem

The allure of the argument from underdetermination argument derives from the fact that the problem it is constructed from seems to cry out for explanations that appeals to prior knowledge, and therefore inductive inference of some kind (Hatfield, 2002). The problem, recall, is that sensory inputs are underdetermining of their distal cause, thus some other factors must also contribute to the determination of a stable percept of the world. Yet, it does *not* immediately follow, simply from this description, that these other factors include prior knowledge. For that to be the case one would minimally need to show that underdetermination has diagnostic features for an induction problem: it is diachronic, empirical, and extrapolative. These features are connected. If a problem is synchronic, and only involves representing information from the present environment, then it does not obviously requiring extrapolating from past experiences. In which case, one may then doubt whether explaining the phenomenon will require positing unconscious inductive inference at all.

There is good reason to think the underdetermination problem is synchronic, as illustrated by the common example of "shape from shading" (Ramachandran, 1988). In Fig. 2A, horizontally-aligned linear contrast gradients are enveloped by circular contours. These gradients are ambiguous cues to 3D shape since they can be caused by concave surfaces illuminated from below, convex surfaces illuminated from above, or an infinity of further illumination and surface shape combinations (Freeman, 1994; Wagemans et al., 2010). Yet we clearly see those with higher luminance at the top as convex dimples unlike those with higher luminance at the bottom. So the visual system appears to make an "assumption" about the typical direction of surface illumination direction. One possibility is that this assumption is overtly represented and recruited by an inferential process. However, a common alternative explanation is that the visual system may internalize, without representing, environmental regularities via *natural constraints* on its organization.[10] For example, in their classic theory of edge detection Marr and Hildreth (1980) proposed that the retina respects a "spatial coincidence assumption" such that the outputs of different spatial frequency filters with similar receptive fields are combined, since edges tend to cause illumination changes at mul-

[10] See: Barlow (1990), Hochberg (1981), Marr and Hildreth (1980), Shepard (1984), Pylyshyn (1999).

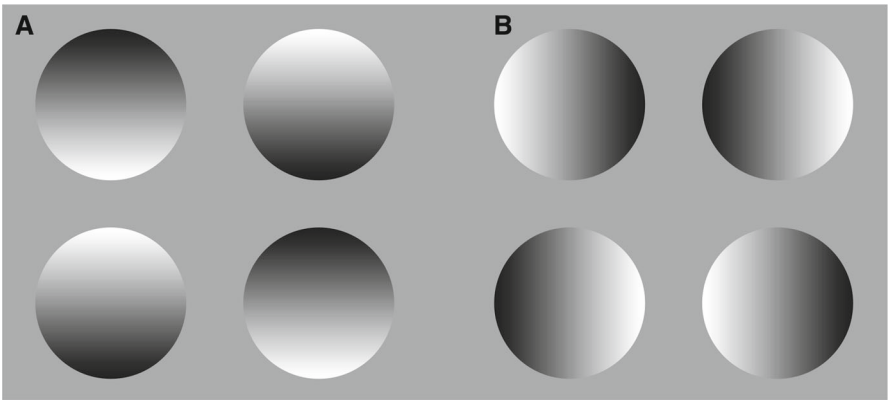

Fig. 2 Circular contours filled with **A** horizontally and **B** vertically aligned linear contrast gradients

tiple spatial frequencies. However, the retina does not *represent* this assumption.[11] Similarly, the assumption that illumination coming from above has been described as natural constraint on how the visual system parses surface shape (e.g. Burge, 2010; Orlandi, 2016).

Adjudicating between these alternative interpretations of the light-from-above assumption requires also taking stock of other facets of the phenomenon of shape from shading. Here are two of them. First, the light-from-above assumption is relatively weak and easily overridden by other cues from shading or shadow (Morgenstern et al., 2011); lighting diffuseness (Morgenstern et al., 2014); the presence of specular highlights (Adams & Elder, 2014); and the shape of the bounding contour (Todorović, 2014). So only in special cases like Fig. 2A does the assumption appear to play an outsized role (Wagemans et al., 2010). Second, even in these special cases non-visual cues are still essential for determining which direction is "above". In particular, the assumption is not constant when the body is rotated so that the gravitational and visual frames of reference are teased apart (Adams, 2008; Barnett-Cowan et al., 2018; Jenkin et al., 2004). To experience the effect of frame of reference for oneself, simply tilt one's head to the left or right until it is horizontal and which stimuli in Fig. 2B appear dimpled will alternate.

The importance of these two facets is that they reveal how shape from shading may be best characterized as a *synchronic* phenomenon in which multiple visual and non-visual cues are combined to guess at the shape of illuminated surfaces. Several constraints no doubt govern how these inputs are combined, but attention to the details of the phenomenon makes the inferential characterization of the light-from-above assumption increasingly untenable. While I am inclined to think this holds, in general, for how the visual system solves all versions of the underdetermination problem (cf. Burge, 2010; Orlandi, 2014), for present purposes what is important is that the *structure* of the underdetermination problem itself is equally compatible with such explanations. If a visual phenomenon has the diagnostic features of an induction problem, it will

[11] For defenses of this interpretation of spatial coincidence assumption from Marr and Hildreth's theory, see Orlandi (2014), Ritchie (2019).

depend less on the fact that there is underdetermination of the input and more so further facets of the phenomenon in question.

3.2 Inferential interpretations of Bayesian models are Underdetermined

The use of Bayesian modeling in vision science is commonly framed as a vindication of the idea that the visual system carries out unconscious inference to solve the underdetermination problem (Rescorla, 2015).[12] Given its popularity, it is worth considering whether this Bayesian variant of the argument better fits the strategy I have proposed.

Bayesian decision theory is a formal framework for modeling decision-making under uncertainty (Berger, 1985). Central to the framework is the notion of subjective probability, or *credence*, which is a quantitative estimate of the degree of belief of an agent. The framework specifies norms for how an agent ought to (optimally) assign credences to hypotheses given the evidence available. The most familiar norm is Bayes' theorem, which expresses the conditional probability $P(h|e)$, or the probability of the hypothesis h being true given the evidence e, as proportional to the unconditional probability of h being true, $P(h)$, and the likelihood of e given the truth of h, or $P(e|h)$. A separate norm is *conditionalization*, which governs how credences should change with new evidence; that is, upon being presented with e we should update $P(h)$ with $P(h|e)$, or replace the *prior* probability of the hypothesis with the *posterior* probability.

Although these norms are distinct (and justified separately), researchers in cognitive science have developed sophisticated models of a wide range of phenomena using both of these norms (Rescorla, 2021), including many visual phenomena (Knill & Richards, 1996; Yuille & Kersten, 2006). Among them is shape-from-shading, where the light-from-above assumption has been formalized as a prior; that is, the credence for the hypothesis that the illumination of a surfaces is directed from above is greater than for other alternative hypotheses about lighting direction. For example, based on behavioral performance across multiple illumination conditions, some studies suggest that the highest credence may actually be for illumination from above-left (Mamassian & Goutcher, 2001; Sun & Perona, 1998).

When made more explicit, the Bayesian variant is grounded in realist interpretations of Bayesian models according to which they are "approximately true" descriptions of the visual system; in other words, visual processing assigns credences to hypotheses in a manner that conforms to the Bayesian norms of reasoning (Rescorla, 2015, 2021). The alternative instrumentalist interpretation treats Bayesian models as merely predictively useful (Block, 2018; Colombo & Seriès, 2012). According to realists, it is the very success of Bayesian modeling that justifies positing credal states in visual processing. Given the strategy I have adopted, these states must be mental representations with the appropriate content if it is to follow that realism about such models warrants positing unconscious inference in the sense I have articulated. It is far from obvious that that is the case, at least without further argument (Orlandi, 2016). How-

[12] Though any connection of Bayesian modeling to the actual work of figures like von Helmholtz is rather tenuous (Westheimer, 2008).

ever, a more fundamental issue is that Bayesian models are not inherently models of information-processing in the first place.

Inferences are a kind of process: we deliberate from certain premises to conclusions, like guessing that the birds in Fig. 1A are ospreys. However, Bayesian models are not necessarily considered process models. Users of the framework are explicit about this (e.g. Griffiths et al., 2010), as Bayesian models are frequently described as a (rational) aspect of Marr's (1982) computational theory, which is a specification of what function a system is trying to carry out, and why (Ritchie, 2019; Shagrir, 2010). As a normative framework, Bayesian modeling provides possible constraints on the problem a system is trying to solve and its ideal solution, but the mapping to the process that solves the problem is many to one (Knill & Richards, 1996; Griffiths et al., 2010; Lake et al., 2017). In this way, Bayesian models *underdetermine* the form of the process that may conform to Bayesian norms. This fact, and the connection to Marr's computational theory, has been used to argue in favor of instrumentalism about Bayesian modeling (Colombo & Seriès, 2012). However, what I think it shows is that what interpretation of Bayesian models we adopt once more depends on the contours of the phenomenon being explained.

This latter point is well illustrated by so-called "rational process models", which specify algorithms that approximate a process that carries out operations over credal states (Griffiths et al., 2015). For example, Shi et al. (2010) used exemplar models of category learning to carry out importance sampling (a form of approximate Bayesian decision-making), where events remembered from the past act as samples from the prior. In their study this approach was applied to psychological tasks, including the number game, where having been told a set of natural numbers fit in a category, participants must guess the probability that a particular number is also included (Tenenbaum & Griffiths, 2001). In this case the model is used to describe an inferential process, but that is because playing the number game, as a phenomenon, requires deliberation.

What then of Bayesian solutions to the underdetermination problem? These are often interpreted in non-inferential terms as reflecting natural constraints (Knill et al., 1996). In the case of shape-from-shading, the light-from-above prior is one of them.[13] To insist otherwise requires some evidence that the phenomenon also has the features of an induction problem. For example, in defense of inferential realism about the light-from-above prior, Rescorla (2015, 2021) points to the fact that the prior can be altered as suggested by the results of Adams et al. (2004). In their study, visual and haptic stimulation was manipulated to suggest a shift in illumination direction resulting in a change in credences, which also impacted performance on judging which side of a bar was lighter. Showing that the visual system can be recalibrated at a time, and that this change is preserved when performing another task, suggests a phenomenon that is diachronic, though without further evidence this is consistent with the hypothesis that natural constraints are flexible. But whether or not Rescorla's argument succeeds note that it has little to do with Bayesian modeling as such, but instead depends on features of the phenomenon beyond the platitude that the sensory input is underdetermined.

[13] Priors being reflected in natural constraints also makes sense of how they might be innate, but not in a way that supports an inferential interpretation of Bayesian models (cf. Scholl, 2005).

In summary, appealing to Bayesian modeling does not avoid the defect in the argument from underdetermination; if anything, it only further emphasizes the flaw: underdetermination does not present an inductive inference problem for the visual system. The foregoing is not a decisive blow against the argument from underdetermination. However, it does suffice for motivating the search for an alternative aspect of visual processing that may involve unconscious inference. The invariance problem offers such an alternative.

4 The argument from invariance

I understand "visual object recognition" as the process of applying mental representations (for a category or individual identity) in the visual system to label the objects that we see (DiCarlo et al., 2012; Riesenhuber & Poggio, 2000). Of course we visually recognize many other things that preoccupy vision scientists including shapes, colors, scenes, and materials. Appropriately adapted, my argument may apply to those cases as well. The argument is also not entirely new. Ibn Al-Haytham took recognition as his paradigm example of unconscious inference (Sabra, 1978), so the argument can also be thought of as a vindication of his theory.[14] In this section I first illustrate the importance of the invariance problem to explaining object recognition and why it is distinct from underdetermination problem. I then detail why the invariance problem has all the diagnostic features of an inductive inference problem and why the general form of the proposed solutions in vision science meet the requirements on representation and process for unconscious inference.

4.1 What is the invariance problem?

The invariance problem is this: we never see objects in the distal world under identical viewing conditions, yet visual perception is largely invariant to identity-preserving transformations of visual input. One may have seen ospreys in the past, but each new time the viewing conditions will be different: the daylight illumination, the viewing distance, orientation, and bodily configuration will all differ. Yet, across these multitude dimensions of change, *what* is seen remains the same, even though *how* it is seen does not.[15] In virtually all discussions of object recognition the invariance problem is the central explanandum:

[14] Buckner (2019b) argues that categorization behavior picks out the the lower bound on rational practical inference. There are commonalities between Buckner's argument and the one present here, as he also acknowledges that it may be grounded in similar claims about theoretical inference (Buckner, 2019b, p. 702). However, a notable difference is that his argument identifies a role for metacognitive feelings in guiding the deliberative process and so does not concern unconscious inference as such.

[15] Object recognition, so characterized, should be distinguished from object *detection*, which concerns whether we see an object, but not what it is. Instances of object (or visual feature) detection are unlikely to involve unconscious inductive inference in the sense I have spelled out if they reflect hardcoded natural constraints that leave no room for learning and generalization—especially when they lead to a reflex-like behavior. For example, "sign stimuli" that cause fixed action plans by organisms involve detection of a target that is innately specified and not open to learning or modulation from experience. Hence, the

> The recognition of visual objects is a fundamental, frequently performed cognitive task with two essential requirements, invariance and specificity. For example, we can recognize a specific face among many, despite changes in viewpoint, scale, illumination or expression. The brain performs this and similar object recognition and detection tasks fast and well. But how? (Riesenhuber & Poggio, 1999, p. 1019)

> Visual object recognition is an extremely difficult computational problem. The core problem is that each object in the world can cast an infinite number of different 2-D images onto the retina as the object's position, pose, lighting, and background vary relative to the viewer …Yet the brain solves this problem effortlessly. (Pinto et al., 2008, p. 1)

> Invariance is a central problem in vision: How do we recognize an object or scene to be the same …across changes in view, size, lighting, configuration, and, even, in the case of category invariance, exemplars? (Gauthier & Tarr, 2016, p. 378)

In so far as object recognition is a fundamental aspect of how we see the world, if the underdetermination problem presents a general challenge for our explanations of visual processing, then the invariance problem does as well (Rust & Stocker, 2010). Both problems relate to how the visual system ultimately generates a stable percept given that there is no one-to-one mapping between visual inputs and their distal causes. In this respect, they can be both thought of as involving sensory uncertainty. However, the problems are conceptually distinct, in at least two respects.

First, they reflect different aspects of visual perception. Solving the underdetermination problem allows us to see a stable and coherent world of, or concerns *seeing*, while solving the invariance problem allows us to make sense of what we see in this world, or concerns *seeing-as*. For example, the former relates to how the visual system arrives at the representation of the object in Fig. 1B that is stable and determinate despite the ambiguity in the input; the latter relates to how we recognize it as an osprey despite never having seen the photo before. Second, they differ in structure. The underdetermination problem relates to how the visual system generates a single stable percept given that any single proximal sensory *input* is compatible with an infinite different distal *causes* (Fig. 3A). In contrast, the invariance problem relates to how we manage to represent the same distal *cause* even though across viewing conditions it can produce a near infinity of different proximal sensory *inputs* (Fig. 3B).

These differences are important for avoiding two possible confusions. The first is that the underdetermination problem is often associated with perceptual constancies, the fact that perception of shape, color, or size tends to be relatively stable across changes in the sensory input (Cohen, 2015). Although "constancy" and "invariance" are sometimes used interchangeably, constancies primarily relate to phenomena of seeing *not* seeing-as. For example, we tend to see the color of surfaces as being the same despite often drastic differences in incidental illumination, however the information-processing challenge presented by color constancies is one of underdetermination:

Footnote 15 continued
processes that control the release of behavior in such cases will not qualify as instances of unconscious inductive inference.

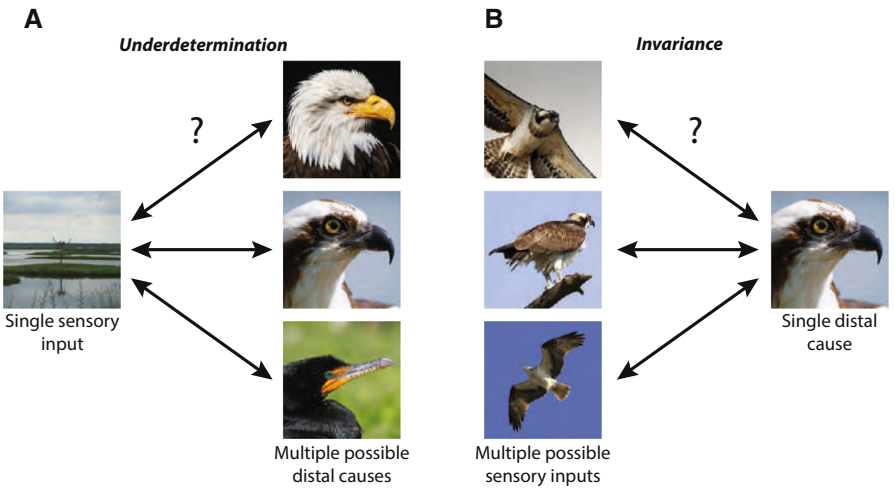

Fig. 3 Two mappings between sensory inputs and their distal causes, after Rust & Stocker (2010, Fig. 2). **A** A single sensory input underdetermined by possible distal causes. **B** a single distal cause that can produce multiple possible sensory inputs

that we are able to discriminate color even though illumination and reflectance are confounded in the input (Foster, 2011).[16] The second possible source of confusion is that underdetermination is also present when we are trying to recognize what we see. For example, because of the importance of spatial frequency tuning to face perception, one stimulus manipulation is to convolve face images filtered at different spatial frequencies with different levels of noise, which allows for parametric variation in the discriminability of face stimuli (e.g. Harmon, 1973; Näsänen, 1999). With such a manipulation, the fact that we see the face images *as* more or less ambiguous (depending on the amount of noise added) entails that the visual system has already made a guess with respect to the ambiguity relevant to the underdetermination problem.

Explanations of visual-processing must ultimately take stock of both the underdetermination and invariance problems (Rust & Stocker, 2010). But it should be clear now that they are distinct challenges, and only the latter reveals a form of unconscious inference in the visual system, as I will now show.

4.2 Solving the invariance problem requires unconscious inference

Following the groundwork laid down earlier, the argument from invariance proceeds in two stages: first, the invariance problem has all the diagnostic features of an induction problem; Second, the type of unconscious process that is posited to explain how the visual system solves the invariance problem exhibits the required commonalities for unconscious inference. Therefore, the fact that an unconscious process in the visual system solves an induction problem gives us good reason to characterize object

[16] Our ability to discriminate colors is also typically considered distinct from the phenomenon of color categorization (see e.g. Witzel & Gegenfurtner, 2018).

recognition as involving a form of unconscious inference. Let us go through each stage in turn.

The form of the invariance problem suggests it has all three diagnostic features for an induction problem. First, the invariance problem is diachronic, since it is defined in terms of how we perceive objects as being the same (in terms of identity or category) across different viewing conditions separated in time. To explain the phenomenon requires accounting for how information from past and present viewings are related to generate a current visual representation of an object (e.g. as an osprey). Second, the problem is empirical, since it concerns information acquired from past visual experience with objects. To explain the phenomenon requires accounting for how this past information is represented and recruited in the present. Third, the problem is extrapolative, since our present viewing conditions are always constitutively different from those of past experiences, even when encountering the same individual object. Thus, perhaps most fundamentally, explaining the phenomenon requires accounting for how we are able to generalize from unlike circumstances across identity-preserving transformations of viewpoint.

That the invariance problem has all three of these diagnostic features is also illustrated by object recognition tasks that focus on how we form representations of novel objects. Here are two classic paradigms. In the first, subjects are presented with different novel stimuli from a restricted set of viewpoints, such as orientations in depth, and they are then tested to see how their performance generalizes to novel viewpoints of the objects not presented during training (Tarr & Pinker, 1989). In the second, subjects are similarly trained on novel objects, "Greebles", but in a way that involves focusing on local diagnostic features before generalizing the knowledge to new individual Greebles, with subsequent stages of the task intended to achieve expert level performance (Gauthier & Tarr, 1997). These two paradigms have been used extensively to investigate (respectively) how the representations we build up to recognize objects are influenced by viewpoint and how visual expertise might help explain how we see familiar categories like faces. For present purposes, what is notable is that they presuppose that object recognition exhibits the three diagnostic features of an induction problem. For it is the very nature of these tasks to investigate how participants generalize from past training experiences to novel test ones.[17] Thus, they also make clear that the invariance problem is a kind of induction problem for the visual system.

Next, consider that all information-processing explanations of object recognition take on the same form: a representation of a perceived object and its visible properties (e.g. an object in the sky) is built up through the processing stages of the visual system and compared to those of individual identities or object categories (e.g. stored representations for different bird species). This *matching process*, as I will call it, occurs automatically and is generally considered inherent to the visual system (DiCarlo et al., 2012; Gauthier & Tarr, 2016; Riesenhuber & Poggio, 2000). The details are often murky as to how the matching process results in a single percept that attributes the relevant label to an object in the environment. So explanations that posit a matching

[17] Of course, "in the wild" object recognition does not involve an explicit partition between training and test experiences with explicit feedback. Some behavioral paradigms also exclude explicit feedback during training, such as those that involve passive viewings of sequential viewpoint images of objects where learning is via temporal association (Cox et al., 2005; Tian & Grill-Spector, 2015; Wallis & Bülthoff, 2001).

process should not be considered a *complete* explanation of object recognition. Still, the matching process at the heart of these explanations has the required commonalities indicative of unconscious inference.

First off, the process includes types of representations with the right contents: one represents information about the object we presently see and its visible properties and the other stored information of the appearance properties for different object identities and categories. In both cases it is also plausible that these are *mental* representation as their content is both distal and robust. First, distalness is often defined *in terms* of invariance as the representation of aspects of our environment that remain the same across, and are distinct from, proximal sensory inputs (Burge, 2010; Mole & Zhao, 2016; Orlandi, 2014). The synchronic representations of an object at a time, under particular viewing conditions, is generally thought to occur at the later stages in the visual processing hierarchy in which increasing levels of specificity and invariance in representational content occur (DiCarlo et al., 2012). So to the extent the greater invariance in content entails greater distality, and object recognition trades in visual representations that exhibit the most wide ranging invariance, the content is distal as well. In turn, our representations of object category or identity are certainly considered distal in so far as they must subserve generalization to novel circumstances. Second, the phenomenon of object recognition *itself* is typically used to illustrate the robustness of content via examples of misrepresentation (Fodor, 1990). Indeed, both types of representations that feature in the matching process have content that appears to be robust in the requisite way: if we mistake a goshawk for an osprey, the categorical representation for the class of osprey is mistakenly tokened, but this could be because of misperception of crucial distinguishing features such as wing shape or plumage.

Next, the matching process is also rule-following and recruits memory. First, in so far as matching is, by hypothesis, a kind of information-processing, it will be rule-following in the minimal sense that computation (in general terms) involves rules defined over representational states of some kind (Piccinini & Scarantino, 2011). As it has just been argued, the states in question are also plausibly mental representations and so the matching process would seem to conform to the idea of a bare inferential transition: given some criteria for what determines a match, if they are satisfied, then the relevant mental label will tend to be applied to the object in the environment that is being represented. Second, the details described so far would suggest that positing some kind of memory store for the mental representations of different labels (for individuals or categories) is unavoidable, for inherent in the idea of the matching process is that such labels exist based on past experience and can be applied in novel input conditions.

All these representational and processing commonalities with inference can be illustrated by considering a "geometric" way of characterizing the matching process in terms of a neural population code, which has become prevalent in visual neuroscience (DiCarlo & Cox, 2007; DiCarlo et al., 2012). Under this construal, the representation of an object at a time, in terms of its visually discernible properties, is encoded as a point in a multi-dimensional visual feature space (as implemented in patterns of neural activity). In turn, representations for category or identity make up distinct regions in this space. The matching process is then a result of applying a decision rule to the new encoding space, in a way similar to machine learning classifiers. For example, a

particular point in the space may have never been tokened before, but if it is located within a region that constitutes the representation of a familiar category (e.g. osprey), the the visual system attributes the property of being an osprey to the object. The process is rule-following and memory-involving because the transition from tokening of a point in the encoding space to the labeling of a stimulus based on the representation that subsumes the point in the space is a reliable, rule-following one, and the encoding space itself is a kind of long-term memory story for representations of previously encountered object types.

This geometrical construal, while increasingly pervasive, is not uncontroversial. In particular, for it to manifest the commonalities with inference in the way just described, then minimally regions of a state space must be possible vehicles for content, and decision-rule operations defined over those representing spaces suffice for a form of bare inferential transition (Gärdenfors, 2004; Shea, 2007).[18] However, this construal suffices to show one way in which the matching process could be realized, granting these auxiliary assumptions. As it happens, this geometric characterization also comports with the theory of Ibn Al-Haytham, who claimed the reason one recognizes the osprey in Fig. 1B is because its visible properties are more similar to ospreys we have seen in the past than other birds. In this way his theory conforms to an intuitive characterization of a "nearest neighbor" classifier of a feature space (Pelillo, 2014). So in a theoretically substantive way, the matching process posited by some modern theories of the neural basis of object recognition also conforms closely to the form of unconscious inference first proposed by Ibn Al-Haytham.

To recap, the invariance problem for object recognition exhibits key diagnostic features of an induction problem. Furthermore, although I have only presented the general form of the explanations of object recognition, the matching process central to these explanations exhibits several key commonalities with inductive thinking. Thus, in an explanatorily substantive way, they involve positing a form of unconscious inference.

5 Challenging the argument

Having built up the argument from invariance, in this section I consider some ways to bring it down. The first concerns the ubiquity of the matching process in explanations of object recognition; the second concerns whether the requirements for unconscious inference that I have identified have indeed been satisfied; and the third concerns the status of object recognition as a perceptual phenomenon.[19]

[18] If the state space is encoded in distributed patterns of neural activity (say) then the information-processing rules will also be defined with respect to the sub-symbols that make up the pattern, rather than the dimensions of the state space themselves. Thus, it must be further assumed that operations over distributions of sub-symbols is one way in which inferential transitions over state spaces can be implemented.

[19] A response I will not consider is that there is no invariance problem. For example, Gibson (1979), and many following in the ecological perception tradition (e.g. Burton & Turvey, 1990), reject the existence of the invariance problem because they posit a unique mapping between the distal world, proximal stimulation, and perception. However even to some within the ecological psychology tradition he started, the existence of such a "one-to-one-to-one" mapping is empirically untenable (Withagen & Chemero, 2009).

5.1 How ubiquitous is the matching process?

For the argument from invariance to succeed, positing a matching process must be both fundamental and widespread in explanations of object recognition. One may doubt I have provided sufficient evidence of this. Note that it is not enough to offer the platitude that the matching process features in our "best" theories or insist one or two curated studies are representative—as is arguably the case with the Bayesian variant of the argument from underdetermination (e.g. Rescorla, 2015). Instead, to address this concern I briefly review three debates about object recognition that have exercised vision scientists. In each case, the core understanding of the phenomenon, and how it is to be explained in terms of a matching process, is largely agreed upon.

The first debate concerns the format of representations for 3D shape recognition. Early theories posited viewpoint independent structural descriptions of object shape built from volumetric primitives, which are compared to structural descriptions of objects stored in long-term memory (Biederman, 1987; Marr & Nishihara, 1978). Later, image-based theories were proposed according to which new viewings of objects are compared to stored representations of objects from previously experienced or canonical viewpoints (e.g. Bulthoff & Edelman, 1992; Cutzu & Edelman, 1994). The debate between these theories centered on whether, in certain theoretically relevant conditions, recognition performance in fact varied with viewpoint (Biederman & Gerhardstein, 1993; Hayward & Tarr, 1997; Tarr & Bülthoff, 1995). Although the debate has fizzled (Hayward, 2003; Stankiewicz, 2003), regardless of whether the representations recruited during object recognition are structural descriptions or image-based, in both cases a form of matching process is a fundamental posit of the two types of theories, since the debate concerns the *format* of the representations that are being matched.

The second debate concerns the domain-specificity of category-selective areas in human visual cortex. One of the first areas discovered was the fusiform face area (FFA) based on increase in fMRI BOLD signal amplitude in response to faces relative to other stimuli like scenes (Kanwisher et al., 1997; McCarthy et al., 1997). Some studies found that visual expertise for novel (e.g. Greebles) or familiar (e.g. birds or cars) object categories also induced greater BOLD responses in FFA (Gauthier et al., 1999, 2000; Xu, 2005). Thus, debate centered on whether FFA has a domain-specific specialization for representing faces or a domain-general specialization for visual expertise (Bukach et al., 2006; Kanwisher & Yovel, 2006). At present, interest in the expertise hypothesis as a rival explanation has somewhat dissipated, and FFA is typically identified as part of a larger network of cortical face areas (for a review, see Duchaine & Yovel, 2015). However, under either hypothesis, FFA is assumed to play some role in building up representations for the matching process at the heart of recognition. At issue is whether this role is exclusive to representing faces or other expertly learned categories as well.

The last debate concerns the use of deep neural networks (DNNs) as models of visual processing.[20] The interest in DNNs initially came from their (near) human-like

[20] For a philosophical introduction to DNNs, see Buckner (2019a). Briefly, architecturally what distinguishes DNNs from earlier generations of neural networks is the following: first, they are "deep" in

performance on image classification tasks (Krizhevsky et al., 2012; LeCun et al., 2015), and the apparent similarity between the representations in layers of DNNs trained on these tasks and neural activity in category-selective visual cortex of primates (Cadieu et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014). Based on such findings, DNNs continue to be used as models of visual processing, and object recognition in particular (Kriegeskorte, 2015; Lindsay, 2020). However, many findings suggest tempering such enthusiasm (Serre, 2019). Classification performance can be disrupted by "adversarial" examples that fool DNNs into incorrectly labeling images that look nothing like the assigned category (Brendel et al., 2017; Goodfellow et al., 2014), while the extent of the correspondence between network representations and the brain is a source of ongoing study (Rajalingham et al., 2018; Xu & Vaziri-Pashkam, 2021). Despite ongoing disagreement about how the networks should be interpreted or utilized (Saxe et al., 2020), there is agreement about the form of the underlying matching process as one that requires comparing incoming signals to stored category labels.

Taken as a whole, then, all three of these debates provide strong evidence that the matching process is a fundamental and widespread posit of research on object recognition. To the extent that the matching process satisfies the conditions for unconscious inference that have been articulated, it follows that a form of unconscious inference is similarly a ubiquitous posit in explanations of the phenomenon.

5.2 What does unconscious inference require?

Another way to challenge the argument from invariance is to raise doubts that the requirements for unconscious inference that I laid out are in fact satisfied by the matching process at the heart of explanations of object recognition. I will consider two objections of this sort.

The first relates to rule-following, which some proponents of unconscious inference claim requires representation of inferential rules themselves (e.g. Rock, 1983). Many have questioned whether visual processing involves rule-representation of this form. As mentioned earlier, a common explanatory approach is to posit natural constraints, which preclude the need for positing rule-representations, and it has been claimed, unconscious inference as well (Burge, 2010; Orlandi, 2014). For example, since neural networks do not overtly represent rules and yet approximate visual processing, some have claimed they therefore undermine the case for unconscious inference (Hatfield, 2002, p. 136; Orlandi, 2014, pp. 46–49). Thus, if rule-representation is also necessary for rule-following then my argument is at best incomplete, and at worst, demolished.

I have three replies. First, as many have pointed out, rule-representation cannot be a necessary condition for *all* inferences on pain of infinite regress (Boghossian, 2014; Carroll, 1895; Fodor, 1987; Quilty-Dunn & Mandelbaum, 2018). Briefly: if inference

Footnote 20 continued
the sense that they have more than one hidden layer (sometimes even hundreds of them). Second, they involve a mixture of different kinds of layers, such as convolutional and fully connected layers. And third, they are sparsely connected. For example, convolutional layers may only be connected with a subset of nodes in the next layer. Technologically, the initial critical advance was to leverage GPUs to train networks with several convolutional layers on complex stimulus sets using error back propagation, which had not previously been feasible (Krizhevsky et al., 2012).

always requires representing a rule linking premises (e.g. modus ponens), then in order to follow that rule, a second-order rule is required that references the first rule, but that second-order rule must itself feature in a third-order rule, and so on. So even when it comes to the sort of deliberation that typifies cognition, rule-following must exclude rule-representation at some level. Second, rule-representation is plausibly connected to the idea that, when deliberating, we consciously "take" the premises to support the conclusion (Boghossian, 2014). In so far as unconscious inference, by definition, precludes such awareness, then rule-representation is ill-motivated as a requirement.[21] Third, the idea of natural constraints in the visual system is wholly compatible with the idea of bare inferential transitions that I have been relying on (Quilty-Dunn & Mandelbaum, 2018; Wright, 2014). Indeed, natural constraints may provide an *articulation* of how the visual system could carry out such transitions.[22] For these reasons, the argument from invariance is not beholden to the requirement that rule-following presupposes rule-representation.

The second objection relates to the conditions for mental representations. Earlier I argued, on general grounds, that the matching process compares forms of mental representation. In order to show that this process is fundamental and pervasive, above I pointed to several debates about object recognition. However, I did not provide evidence that these more specific lines of research posit internal states with content that is both distal and robust. In fact, there is good reason to think they may not. For example, one may reasonably doubt that image-based theories of 3D shape recognition posit states that represent the distal environment as opposed to proximal image features. Furthermore, even if they do posit mental representations, then one would then still need to show that they satisfy Ramsey's (2007) "job description challenge" for mental representation: that it is in virtue of the posited states being mental representations that they are able to play their explanatory roles. Thus, I have not yet shown that, across these debates, positing mental representations is either widespread, or necessary, to the explanations that have been pro-offered.

In reply I would distinguish between two issues. The first is whether positing mental representations is necessary for the explanation of a visual phenomenon. I have already made arguments that this is the case. The second is the role of mental representation in interpreting *particular* theories or models of the phenomenon, given a prior commitment to the explanations requiring mental representation. Regarding this issue, it is important to keep in mind that the theories and models I canvassed may either be incomplete, abstract from important details, or simply be incorrect, but none of these alternatives would, by themselves, give us reason to doubt that object recognition involves mental representation. For example, image-based theories, as *models* of 3D shape recognition, may reflect the incomplete and abstract form of our theorizing. Thus, the fact that I have not shown that mental representations are ubiquitously

[21] The same is true if the taking condition is characterized as a consciously available evaluative valence (Buckner, 2019b; Carruthers & Ritchie, 2012).

[22] Note that this is consistent with the earlier claim that priors being reflected in natural constraints undermines the underdetermination argument. Under the characterization I have offered of induction problems and their solutions, (i) the inputs must be overtly represented, but (ii) not the transition rules that govern the relationship between them. Priors as natural constraints is inconsistent with (i), but inferential transitions as natural constraints is consistent with (ii).

explicitly posited, or that the job description challenge is consistently met, does not undermine the argument.

5.3 Is object recognition really perceptual?

At this point, one may agree that object recognition involves unconscious inference, but question whether the conclusion is interesting. Object recognition is, after all, part of what is sometimes called visual *cognition*. So it should come as little surprise that seeing an osprey as such involves unconscious inference as it reflects us learning and remembering from experience (Hatfield, 2002). How much standing we should give this concern depends in part on the extent to which object recognition should be considered a perceptual phenomenon in the first place. For example Thomas Reid, who perhaps first clearly distinguished sensation and perception, considered forms of acquired perception, like object recognition, just as perceptual as other aspects of seeing (Copenhaver, 2010). Still, so far I have assumed that object recognition is principally a perceptual phenomenon, without argument. Furthermore, object recognition has increasingly featured as a test case for determining how perception and cognition should be distinguished (Firestone & Scholl, 2016; Mandelbaum, 2018). While I do not intend to take a stand on that debate here, below I offer three arguments in favor of the claim that object recognition is perceptual, before addressing some reasons one might reject it.

Why think object recognition is indeed perceptual? First, prima facie the experience of object recognition appears perceptual, as illustrated by classic demonstrations of the contrast between seeing and seeing-as. In two-tone "Mooney" images once we parse the image we do not simply see a coherent scene we see the central object *as* a dalmatian (Fig. 4A). When we look at the "do-it-youself" object of Biederman (1987) we clearly see an object but what is *missing* is that we see it as anything in particular (Fig. 4B). For bistable images like Rubin's vase (Rubin, 1915) beyond the figure-ground reversal what objects we see also switches (Fig. 4C). A similar effect occurs when we look at the infamous Duck-Rabbit popularized by Wittgenstein (1953), except there is no change in figure-ground assignment (Fig. 4D). Denying that object recognition is perceptual requires explaining away these experiences, rather than simply taking them at face value.

Second, a key feature of object recognition is that it often occurs quickly—∼120 ms, or about as fast as it takes for visual signals to reach category-selective areas of visual cortex (DiCarlo et al., 2012). This point is illustrated by human and primate studies using time-series (cellular recordings or M/EEG) decoding methods in which information about stimulus category latent in neural signal patterns reach peak discriminability 150 ms post-stimulus onset (Carlson et al., 2013; Hung et al., 2005; Isik et al., 2014). While decoding results do not license direct conclusions about representational content being reflected in neural signals (Ritchie et al., 2019), these results are just a sample of many lines of converging evidence that suggest object recognition is often subserved by a rapid feedforward pass of information through the visual system.

Finally, the perceptual status of object recognition depends in part on how one delineates perception from cognition (if at all). As with the notion of inference, there are

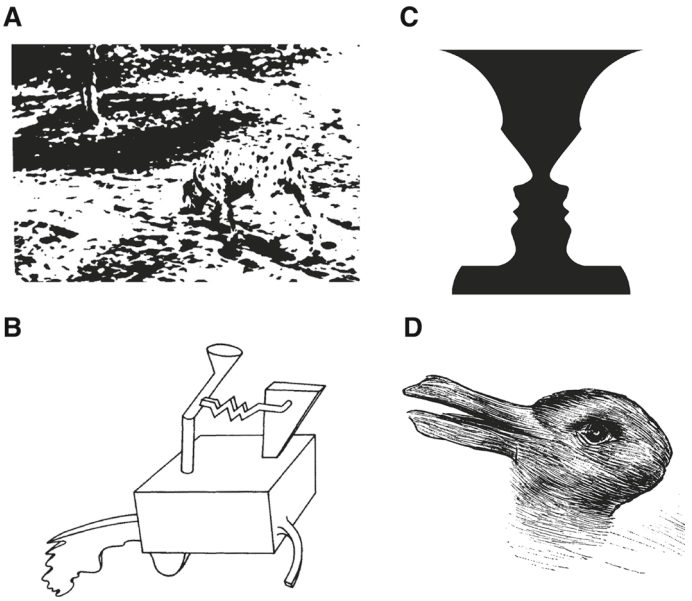

Fig. 4 Four illustrations of the perceptual nature of object recognition. **A** Two-tone "Mooney" image. **B** Biederman's "do-it-yourself" object. **C** Rubin's vase. **D** Wittgenstein's Duck-Rabbit

multiple measures by which we might draw the line. Still, one proposal is that mental representations are perceptual when they are *stimulus-dependent* in the sense that they have the function of being causally sustained when the visual system is presented with proximal sensory inputs (Beck, 2018). Such a condition helps to distinguish how we represent Fig. 1A and B differently. In both cases the representations have demonstrative and attribute elements: we attribute the property of being an osprey to *that* thing before us. But only in the case of Fig. 1B is the attributive element stimulus-dependent because we are also attributing the *appearance* of an osprey, which will be constrained by some particular sensory inputs. In contrast, in Fig. 1A we have a perceptually-determined demonstrative thought that we are looking at ospreys, and may be correct in this belief, but the objects on the nest do not *look* like ospreys, when seen from afar away. So by one plausible measure, object recognition is clearly perceptual.[23]

[23] Cermeño-Aínsa (2021) rejects Beck's stimulus-dependence condition based in part on visual categorization as a case study. However, his critique rests on two mistaken claims about visual categorization and how it is explained. The first is that the neural basis of categorization is not specific to visual cortex (Cermeño-Aínsa, 2021, p. 13). This claim runs counter to the vast majority of research in visual neuroscience (DiCarlo et al., 2012). The second is to not properly distinguish between cases like Fig. 1A, B. Cermeño-Aínsa (2021, p. 14) claims that visual categorization is not perceptually grounded because, on the one hand, we can visually categorize without seeing all the distinctive properties of an object so it is not proximally constrained; and on the other, that visual categories involves our conceptual capacities. However, Beck precludes cases like Fig. 1A as perceptual because in such a case *no* diagnostic visual properties of ospreys themselves are visible; being proximally constrained only requires that some of these properties are visible. Furthermore, as also pointed out in the text, attributing *appearances* does not require conceptual capacities.

What about reasons for *denying* that object recognition is perceptual? Let us consider two of them. First, if categorization suffices for a form of *conceptualization*, then if object recognition is perceptual it follows that perception recruits concepts (Mandelbaum, 2018; Prinz, 2002). Many have rejected this claim on the grounds that perception excludes seeing-as (Block, 2014; Burge, 2010). One way to frame this dispute is in terms of the matching process itself and whether it is carried out by the visual system or not; if it is, then perception involves conceptualization; if not, then object recognition is not perceptual (Mandelbaum, 2018). I reject the first entailment. It is far from clear that representations of the *appearance* of a type of object, which are the type of representations stored and recruited during recognition, qualify as concepts. For example, even if regions of a neural encoding space can be vehicles for mental representation, and the matching process follows the geometric characterization described earlier, it is not obvious that representations in a space that encodes information about object appearances are concepts (cf. Gauker, 2017). Thus, without further premises, my argument is consistent with the possibility of non-conceptual seeing-as.

Second, if memory is considered inherently cognitive, then object recognition is not perceptual. Firestone and Scholl (2016) seem to make this claim to address results of studies that purport to show that perception can be cognitively penetrated. For example, the fact that the memory component of recognition can be influenced by information that an observer is consciously aware of, such as the name of a word making it easier for a stimulus to break through to awareness during continuous flash suppression, has been interpreted as showing that perception is cognitively penetrable (Lupyan & Ward, 2013). However, such results at most show that perception is not informationally encapsulated, which need not be be treated as a requirement for distinguishing perception and cognition (Beck, 2018; Ogilvie & Carruthers, 2016).[24] For object recognition to be cognitively mediated in a substantive sense would presumably require some level of cognitive *control* on the matching process itself; it would require a capacity to override the strong stimulus-dependence that determines our representation of the osprey in Fig. 1B, by convincing ourselves we are looking at (say) an eagle instead. So, priming effects on recognition can be dismissed as evidence of cognitive penetration without also denying that recognition is wholly a perceptual phenomenon.

6 Conclusion

The idea that visual information-processing involves unconscious inference is one of the theoretical pillars for much of vision science. I have attempted to provide a novel basis of support for this pillar. We began the present discussion by laying groundwork for an argumentative strategy focused on whether positing a form of unconscious inference plays a role in explaining particular aspects of visual processing. This was used to evaluate the most influential argument in favor of unconscious inference, the argument from underdetermination. This well-known argument, even under a Bayesian

[24] Another consideration is that evidence of cognitive penetration may even be compatible with (or even provide evidence in favor of) information encapsulation, despite the common assumption to the contrary (Clarke, 2020).

guise, is ill-fit to the proposed strategy. In its place, an alternative argument centered on the invariance problem for visual object recognition was constructed. As I have shown, explaining how the visual system overcomes the invariance problem reveals important commonalities between perception and thought. Identifying these commonalities help us recognize why vision is inferential.[25]

Author Contributions JBR solely contributed to this work.

Declarations

Conflict of interest The author declares that there is no conflict of interest.

References

Adams, W. J. (2008). Frames of reference for the light-from-above prior in visual search and shape judgements. *Cognition, 107*(1), 137–150.

Adams, W. J., & Elder, J. H. (2014). Effects of specular highlights on perceived surface convexity. *PLoS Computational Biology, 10*(5), e1003576.

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7*(10), 1057–1058.

Aggelopoulos, N. C. (2015). Perceptual inference. *Neuroscience and Biobehavioral Reviews, 55,* 375–392.

Barlow, H. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research, 30*(11), 1561–1571.

Barnett-Cowan, M., Ernst, M. O., & Bülthoff, H. H. (2018). Gravity-dependent change in the 'light-from-above' prior. *Scientific Reports, 8*(1), 1–6.

Beck, J. (2018). Marking the perception-cognition boundary: The criterion of stimulus-dependence. *Australasian Journal of Philosophy, 96*(2), 319–334.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–117.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*(6), 1162–1182.

Block, N. (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research, 89*(3), 560–572.

Block, N. (2018). If perception is probabilistic, why does it not seem probabilistic? *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1755), 20170341.

Boghossian, P. (2014). What is inference? *Philosophical Studies, 169*(1), 1–18.

Brendel, W., Rauber, J., & Bethge, M. (2017). Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. arXiv:1712.04248

Buckner, C. (2019a). Deep learning: A philosophical introduction. *Philosophy Compass, 14*(10), e12625.

Buckner, C. (2019b). Rational inference: The lowest bounds. *Philosophy and Phenomenological Research, 98,* 1–28.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences, 10*(4), 159–166.

[25] Thank you to Bence Nanay, Bryce Huebner, Cameron Buckner, David Barack, Evan Westra, and the anonymous referees of this journal, for their helpful feedback on earlier versions of this manuscript. This work was also previously presented at Johns Hopkins University. Thank you to the audience there, and in particular, Chaz Firestone, Jorge Morales, and Steve Gross, for their feedback on the project. This research was supported by the Intramural Research Program of the National Institute of Mental Health (ZIAMH002909 awarded to Chris I. Baker).

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, 89*(1), 60–64.
Burge, T. (2010). *Origins of Objectivity*. Oxford University Press.
Burton, G., & Turvey, M. T. (1990). Perceiving the lengths of rods that are held but not wielded. *Ecological Psychology, 2*(4), 295–324.
Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., et al. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Computational Biology, 10*(12), e1003963.
Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision, 13*(10), 1–1.
Carroll, L. (1895). What the tortoise said to Achilles. *Mind, 4*(14), 278–280.
Carruthers, P., & Ritchie, J. B. (2012). The emergence of metacognition: Affect and uncertainty in animals. In M. J. In, J. Beran, J. Brandl, & J. P. Perner (Eds.), *Foundations of metacognition* (pp. 76–93). Oxford University Press.
Cermeño-Aínsa, S. (2021). Is perception stimulus-dependent? *Review of Philosophy and Psychology*, 1–20.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181–204.
Clarke, S. (2020). Cognitive penetration and informational encapsulation: Have we been failing the module? *Philosophical Studies, 178,* 1–22.
Cohen, J. (2015). Perceptual constancy. In M. Matthen (Ed.), *The Oxford handbook of philosophy of perception* (pp. 621–639). Oxford University Press.
Colombo, M., & Seriès, P. (2012). Bayes in the brain-on Bayesian modelling in neuroscience. *The British Journal for the Philosophy of Science, 63*(3), 697–723.
Copenhaver, R. (2010). Thomas Reid on acquired perception. *Pacific Philosophical Quarterly, 91*(3), 285–312.
Cox, D.D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). 'Breaking'position-invariant object recognition. *Nature Neuroscience, 8*(9), 1145–1147.
Cutzu, F., & Edelman, S. (1994). Canonical views in object representation and recognition. *Vision Research, 34*(22), 3037–3056.
DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11*(8), 333–341.
DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415–434.
Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science, 1,* 393–416.
Epstein, W. (1973). The process of 'taking-into-account' in visual perception. *Perception, 2*(3), 267–285.
Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39*.
Fodor, J., & Pylyshyn, Z. (1981). How direct is visual perception?: Some reflections on Gibson's "ecological approach." *Cognition,9*(2), 139–196.
Fodor, J. A. (1987). *Psychosemantics*. MIT Press.
Fodor, J. A. (1990). *A theory of content and other essays*. The MIT Press.
Foster, D. H. (2011). Color constancy. *Vision Research, 51*(7), 674–700.
Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature, 368*(6471), 542–545.
Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.
Gauker, C. (2017). Three kinds of nonconceptual seeing-as. *Review of Philosophy and Psychology, 8*(4), 763–779.
Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191–197.
Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673–1682.
Gauthier, I., & Tarr, M. J. (2016). Visual object recognition: Do we (finally) know more now than we did? *Annual Review of Vision Science, 2,* 377–396.
Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature neuroscience,2*(6), 568–573.

Gibson, J. J. (1979). *The ecological approach to visual perception (classic)*. Psychology Press.
Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese, 193*(2), 559–582.
Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. arXiv:1406.2661
Gregory, R. L. (1970). *The intelligent eye*. McGraw-Hill.
Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories*. University of Chicago Press.
Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences, 14*(8), 357–364.
Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science, 7*(2), 217–229.
Harman, G. (1986). *Change in view: Principles of reasoning*. The MIT Press.
Harmon, L. D. (1973). The recognition of faces. *Scientific American, 229*(5), 70–83.
Hatfield, G. (2002). Perception as unconscious inference. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world* (pp. 115–143). Wiley.
Hayward, W. G. (2003). After the viewpoint debate: Where next in object recognition? *Trends in Cognitive Sciences, 7*(10), 1–3.
Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*(5), 1511.
Hochberg, J. (1981). On cognition in perception: Perceptual coupling and unconscious inference. *Cognition, 10*(1–3), 127–134.
Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
Howard, I. P. (1996). Alhazen's neglected discoveries of visual phenomena. *Perception, 25*(10), 1203–1217.
Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310*(5749), 863–866.
Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology, 111*(1), 91–102.
Jenkin, H. L., Jenkin, M. R., Dyde, R. T., & Harris, L. R. (2004). Shape-from-shading depends on visual, gravitational, and body-orientation cues. *Perception, 33*(12), 1453–1461.
Kanizsa, G. (1985). Seeing and thinking. *Acta Psychologica, 59*(1), 23–33.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302–4311.
Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2109–2128.
Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Computational Biology, 10*(11), e1003915.
Kiefer, A. (2017). Literal perceptual inference. In T. Metzinger & W. Weise (Eds.), *Philosophy and predictive processing* (pp. 257–275). MIND Group.
Knill, D. C., Kersten, D., & Yuille, A. (1996). Introduction: A Bayesian formulation of visual perception. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 1–21). Cambridge University Press.
Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge University Press.
Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci., 1,* 417–446.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst., 25,* 1097–1105.
Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40,* e253.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
Lindsay, G. W. (2020). Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of Cognitive Neuroscience, 33,* 1–15.
Linquist, S. (2018). The conceptual critique of innateness. *Philosophy Compass, 13*(5), e12492.
Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, 110*(35), 14196–14201.
Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition, 81*(1), B1–B9.

Mandelbaum, E. (2018). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research, 97*(2), 267–283.
Margolis, E., & Laurence, S. (2013). In defense of nativism. *Philosophical Studies, 165*(2), 693–718.
Marr, D. (1982). *Vision*. Freeman and Company.
Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society B: Biological Sciences, 207*(1167), 187–217.
Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 200*(1140), 269–294.
McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9*(5), 605–610.
Mole, C., & Zhao, J. (2016). Vision and abstraction: an empirical refutation of Nico Orlandi's non-cognitivism. *Philosophical Psychology, 29*(3), 365–373.
Morgenstern, Y., Geisler, W. S., & Murray, R. F. (2014). Human vision is attuned to the diffuseness of natural light. *Journal of Vision, 14*(9), 15–15.
Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. *Proceedings of the National Academy of Sciences, 108*(30), 12551–12553.
Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research, 39*(23), 3824–3833.
Ogilvie, R., & Carruthers, P. (2016). Opening up vision: The case against encapsulation. *Review of Philosophy and Psychology, 7*(4), 721–742.
Orlandi, N. (2014). *The innocent eye*. Oxford University Press.
Orlandi, N. (2016). Bayesian perception is ecological perception. *Philosophical Topics, 44*(2), 327–351.
Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT Press.
Pelillo, M. (2014). Alhazen and the nearest neighbor rule. *Pattern Recognition Letters, 38*(1), 34–37.
Piccinini, G., & Scarantino, A. (2010). Computation vs. information processing: Why their difference matters to cognitive science. *Studies in History and Philosophy of Science Part A*, 41(3):237–246.
Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics, 37*(1), 1–38.
Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology, 4*(1), e27.
Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. MIT press.
Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*(3), 341–365.
Quilty-Dunn, J., & Mandelbaum, E. (2018). Inferential transitions. *Australasian Journal of Philosophy, 96*(3), 532–547.
Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience, 38*(33), 7255–7269.
Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*(6152), 163–166.
Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge University Press.
Rescorla, M. (2015). Bayesian perceptual psychology. In M. Matthen (Ed.), *The Oxford handbook of the philosophy of perception*. Oxford University Press.
Rescorla, M. (2021). Bayesian modeling of the mind: From norms to neurons. *Wiley Interdisciplinary Reviews: Cognitive Science, 12*(1), e1540.
Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025.
Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3*(Suppl), 1199–1204.
Ritchie, J. B. (2019). The content of Marr's information-processing framework. *Philosophical Psychology,32*(7), 1078–1099.
Ritchie, J. B. (2020). What's wrong with the minimal conception of innateness in cognitive science? *Synthese, 199,* 1–18.
Ritchie, J. B., Kaplan, D. M., & Klein, C. (2019). Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. *The British Journal for the Philosophy of Science, 70*(2), 581–607.
Rock, I. (1983). *The logic of perception*. MIT Press.

Rubin, E. (1915). *Visuell wahrgenommene figuren*. Gyldenalske Boghandel.
Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: Two fundamental challenges for visual processing. *Current Opinion in Neurobiology, 20*(3), 382–388.
Sabra, A. I. (1978). Sensation and inference in Alhazen's theory of visual perception. *Studies in Perception: Interrelations in the History of Philosophy and Science*, 160–185.
Samuels, R. (2002). Nativism in cognitive science. *Mind & Language, 17*(3), 233–265.
Samuels, R. (2004). Innateness in cognitive science. *Trends in Cognitive Sciences, 8*(3), 136–141.
Saxe, A., Nelli, S., & Summerfield, C. (2020). If deep learning is the answer, what is the question? *Nature Reviews Neuroscience*, 1–13.
Scholl, B. J. (2005). Innateness and (Bayesian) visual perception: Reconciling nativism and development. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 34–52). Oxford University Press.
Searle, J. R. (1983). *Intentionality*. De Gruyter Mouton.
Serre, T. (2019). Deep learning: The good, the bad, and the ugly. *Annual Review of Vision Science, 5,* 399–426.
Shagrir, O. (2010). Marr on computational-level theories. *Philosophy of Science, 77*(4), 477–500.
Shea, N. (2007). Content and its vehicles in connectionist systems. *Mind & Language, 22*(3), 246–269.
Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91*(4), 417.
Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review, 17*(4), 443–464.
Stankiewicz, B. J. (2003). Just another view. *Trends in Cognitive Sciences, 7*(12), 526.
Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience, 1*(3), 183–184.
Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1494–1505.
Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*(2), 233–282.
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 629.
Tian, M., & Grill-Spector, K. (2015). Spatiotemporal information during unsupervised learning enhances viewpoint invariant object recognition. *Journal of Vision, 15*(6), 7–7.
Todorović, D. (2014). How shape from contours affects shape from shading. *Vision Research, 103,* 1–10.
Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik: mit 213 in den Text eingedruckten Holzschnitten und 11 Tafeln*, vol. 9. Voss.
Wagemans, J., Van Doorn, A. J., & Koenderink, J. J. (2010). The shading cue in context. *i-Perception, 1*(3), 159–177.
Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, 98*(8), 4800–4804.
Westheimer, G. (2008). Was Helmholtz a Bayesian? *Perception, 37*(5), 642–650.
Withagen, R., & Chemero, A. (2009). Naturalizing perception: Developing the Gibsonian approach to perception along evolutionary lines. *Theory & Psychology, 19*(3), 363–389.
Wittgenstein, L. (1953). *Philosophical investigations*. Wiley.
Witzel, C., & Gegenfurtner, K. R. (2018). Color perception: Objects, constancy, and categories. *Annual Review of Vision Science, 4,* 475–499.
Wright, C. (2014). Comment on Paul Boghossian, "What is inference." *Philosophical Studies, 169*(1), 27–37.
Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15*(8), 1234–1242.
Xu, Y., & Vaziri-Pashkam, M. (2021). Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature Communications, 12*(1), 1–16.
Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences, 10*(7), 301–308.