
Article:
Allen, Keith, Fischer, Eugen and Engelhardt, Paul (2026) Direct or Indirect Realism? Assessing Conflicting Folk Conceptions of Vision. Synthese. 185. ISSN: 0023-7857

Synthese (2026) 207:185
https://doi.org/10.1007/s11229-026-05575-0

ORIGINAL RESEARCH

Direct or indirect realism? Assessing conflicting folk conceptions of vision

Eugen Fischer[1] · **Keith Allen**[2] · **Paul E. Engelhardt**[3]

Received: 18 September 2025 / Accepted: 3 April 2026

Abstract

Longstanding philosophical debates about the nature of perception revolve around a clash between Direct Realist and Indirect Realist conceptions of vision. These deny and affirm, respectively, that vision involves awareness of mental images which represent the physical objects of sight. The assumption that 'the' common-sense conception of vision is Direct Realist shapes these debates. Against this key assumption, recent studies in experimental philosophy have provided first evidence that laypeople are, collectively and often individually, torn between Direct Realist and Indirect Realist conceptions, and regard both as naïve. These findings raise a general methodological issue: Does agreement with two such patently incompatible conceptions reflect genuine beliefs? This paper crucially extends first findings, examines whether agreement ratings are indicative of beliefs, and offers an explanation of how the conflicting beliefs involved could be acquired and maintained. Three studies develop and validate the new *Direct/Indirect Realist Belief Inventory (DIRBI)* which uses agreement ratings with verbal and pictorial stimuli to assess the prevalence of the two conflicting conceptions among laypeople. Findings confirm that laypeople hold relevant beliefs and are, collectively and individually, torn between the two conceptions. We suggest these beliefs are ultimately anchored in ubiquitous implicit knowledge structures, namely, experiential event knowledge about vision and an implicit model of endogenous attention. We conclude that there is no such thing as 'the' common-sense conception of vision that could enjoy the epistemic default status accorded to common sense in many philosophical debates.

Keywords Naïve realism · Direct realism · Indirect realism · Folk conceptions of vision · Belief fragmentation · Belief inventory validation · Experimental philosophy

Extended author information available on the last page of the article

Published online: 21 April 2026

1 Introduction

Common sense combines a chequered reputation with extensive philosophical influence. Albert Einstein famously characterised common sense as 'a deposit of prejudices laid down in the mind before you reach eighteen'.[1] In epistemology, few positions are less popular than default-and-challenge models of justification that accord common-sense beliefs an *epistemic default status* – i.e., suggest we should accept them absent good reasons to the contrary (Brandom, 1994; Williams, 2001). Even so, reliance on common sense is built into philosophical methods ranging from Aristotle's 'endoxic' method (Aristotle, n.d.; cf.; Davia, 2017) to the still popular quest for 'reflective equilibrium' (review: Cath, 2016) and 'material' uses of the 'method of cases' (review: Machery, 2017). Common-sense conceptions of moral *and* natural phenomena have been accorded epistemic default status in particular in debates about 'aporetic problems' (from the Greek 'aporia' or 'impasse') which arise from perceived tensions or conflicts between independently plausible judgments or beliefs (Rescher, 2001). Examples include sceptical problems (review: McKinsey, 2024) and the problems of free will (review: O'Connor & Franklin, 2022) and perception (review: Crane & French, 2021).

Philosophical debates about aporetic problems tend to take for granted that there is such a thing as 'the' common-sense conception, say, of knowledge, agency, or vision.[2] Crucially, these debates also accord this common-sense conception epistemic default status. Thus, many contributions to the experimental philosophy of free will have probed lay conceptions with a view to assessing the burdens of proof for determinist vs indeterminist conceptions of free will – placing the onus of proof on the conception with the less good claim to common-sense status (review: Sommers, 2010). Similarly, many contributors to debates about the problem of perception think that Direct Realist theories which supposedly capture the common-sense conception of vision should be accepted in the absence of good reasons to the contrary (e.g., Allen, 2020; Genone, 2016; Martin, 2002; Reid, 1785/1969; Searle, 2015; Strawson, 1979), and that Indirect Realist theories which allegedly contradict common sense should be maintained only together with an error theory that explains how common sense could go wrong (e.g., Hume, 1739-40/2011; Russell, 1912/2000; Boghossian & Velleman, 1989).

Recent work has challenged the empirical assumptions that underlie this acceptance of common sense as an epistemic default: Fragmentation accounts of belief storage (Bendaña & Mandelbaum, 2021; Sommer et al., 2023) provide a conceptual framework that allows us to understand how mutually incompatible sets of beliefs may be systematically acquired and stored in memory, without updating (see below, Sect. 1.1). Across a wide range of domains, studies from experimental philosophy have documented substantive disagreements in intuitive judgments that reflect conflicts between such competing conceptions within a community and often within

[1] https://www.oxfordreference.com/display/10.1093/acref/9780191843730.001.0001/q-oro-ed5-00003988 Last accessed September 11, 2025.

[2] These philosophical debates are in good company: Psychological research often contrasts scientific theories with 'the' naïve theory of the domain of interest (review: Shtulman, 2017),

the same individuals (review: Knobe, 2026). Two recent studies have followed this up with respect to the competing Direct Realist and Indirect Realist conceptions of vision: An agreement rating task with verbal and pictorial items provided evidence that each conception is endorsed by at least 40% of the lay sample, and that over 20% endorse both (Fischer et al., 2023, with $N = 100$). A scientific accuracy rating study provided evidence that both conceptions are treated as naïve in lay cognition (Fischer et al., 2026, with $N = 200$). Together, these findings suggest that the conflict between Direct Realism and Indirect Realism may be a conflict *within* common sense, i.e., that common sense is itself torn between mutually patently incompatible conceptions of the same phenomenon. If so, there is no such thing as 'the' (unified, coherent) common-sense conception of vision that could enjoy epistemic default status.

The present paper will substantially extend the case for this philosophically significant conclusion by addressing a methodological concern of considerable wider relevance, and by replicating key findings with larger samples and improved materials in the process. The key methodological concern, which calls for methodological innovation, is this: Like many studies in experimental philosophy (including those reviewed by Knobe, 2026) and like the most pertinent studies examining folk beliefs in psychology (review: Winer et al., 2002), Fischer et al. (2023) used agreement ratings. But agreement ratings need not be indicative of standing beliefs or conceptions. This interpretation may be rash, in particular, where items probe views on questions laypeople need not have pondered ever before (e.g., 'Does seeing objects involve seeing mental images?'). Moreover, belief attributions are often thought to be subject to normative principles of charity that constrain the attribution of directly contradictory beliefs to competent thinkers (review: Thagard & Nisbett, 1983). Wherever agreement ratings indicate that a sizeable proportion of participants endorse mutually incompatible conceptions (cf. Knobe, 2026), the apparent *intra*personal disagreement therefore motivates the question of whether the observed agreement ratings are reflective of *beliefs*. This motivation is particularly strong where findings suggest that many participants agreed with both of two competing conceptions *they understood to be incompatible* – as they did in the above-mentioned study on lay conceptions of vision (Fischer et al., 2023).

We therefore now ask two research questions, with an emphasis on the previously neglected first question: Do laypeople hold any Direct Realist or Indirect Realist *beliefs*? More specifically: (RQ1) Are agreement ratings provided in response to statements of Direct Realism and Indirect Realism indicative of such beliefs? And (RQ2): Do laypeople, collectively and individually, hold both Direct Realist and Indirect Realist beliefs?

In psychology, questions about the interpretation of observed responses are systematically addressed by validation studies of the relevant measure. In philosophy, plausible mid-strength principles of charity postulate that we can attribute conflicting beliefs to competent thinkers only if we can empirically explain why they hold such beliefs (Thagard & Nisbett, 1983). Accordingly, the present paper will address our first question (RQ1) in two steps: Through three empirical studies, we will first develop and validate a new belief inventory, the *Direct/Indirect Realist Belief Inventory (DIRBI)* that is specifically designed to capture both interpersonal and intrapersonal belief conflicts (Sect. 2). We will then, second, develop empirical explanations

of why laypeople hold these directly opposing views about vision (Sect. 3). To address our second research question (RQ2) and gauge the extent of inter- and intrapersonal conflict in pertinent beliefs about vision, we will submit data from our validation studies to further analyses (Sect. 2.5). We hope these studies may provide a useful paradigm and inspire efforts to develop and validate further philosophically relevant belief inventories, to study further belief conflicts that give rise to aporetic problems.

We now review the research that motivates the present endeavour, viz., psychological research on naïve theories (Sect. 1.1) and philosophical debates about perception (Sect. 1.2), and then discuss the validation of belief inventories (Sect. 1.3).

1.1 Psychological context: Naïve theories and representational co-existence

Prior to scientific instruction, people form naïve theories of many natural phenomena. These naïve theories are often inconsistent with scientific theories but, even so, persist in the face of scientific instruction (review: Shtulman, 2017). As standardly conceived, *naïve theories* are causal-explanatory structures derived from a combination of innate conceptions, first-hand experience, and social learning (Carey, 2009; Gelman & Noles, 2011; Vosniadou, 1994). Prominent examples include impetus theories of motion (McCloskey, 1983) and substance theories of heat (Reiner et al., 2000; Wiser & Carey, 1983). While they arise for different reasons and take different forms, they tend to give internally coherent, perceptually grounded, object-based explanations of everyday situations that attribute context-invariant properties to objects and agents. By contrast, scientific theories introduce new concepts (like 'force', 'molecule', or 'energy'), postulate imperceptible entities, and focus on process-based explanations at higher levels of abstraction and generalization (Thagard, 2014).

The acquisition of such new concepts and theories does not erase from memory naïve theories that conflict with them. Indeed, conflicting naïve and scientific theories may both continue to compete for use. Evidence is provided by speeded statement verification studies, in which participants assess as quickly as possible the truth of claims that are consistent with both naïve and scientific theories, claims that are inconsistent with both theories, and – crucially – claims that are consistent with one theory, but not the other. Across several different domains from the natural and life sciences, scientifically instructed adults responding under time pressure proved more prone to lapse into endorsing statements in conflict with the scientific theory when these statements were consistent (rather than inconsistent) with the naïve theory, and required more time to correctly assess statements consistent with the scientific theory when these statements were simultaneously at odds with the naïve theory (Potvin & Cyr, 2017; Shtulman & Legare, 2020; Shtulman & Valcarcel, 2012, cf.; Stricker et al., 2021).

Such findings have been regarded as evidence for the *'representational co-existence'* of competing conceptions of the same phenomenon (Barlev et al., 2018): On this interpretation, the representation of the naïve theory is initially activated; where it is in conflict with the subsequently activated scientific theory, correct responses require prior suppression of the prepotent naïve response – an effortful process impeded by time pressure. This interpretation is supported by the exclusion of the

most pertinent confound, viz., low-level associations (Shtulman & Legare, 2020), and by neuroscientific findings: People who endorse scientific claims at odds with naïve theories show higher activity in brain areas associated with error monitoring and inhibitory control (Foisy et al., 2015; Mareschal, 2016; Masson et al., 2014).

Representational co-existence extends beyond science to further domains involving social learning: Speeded statement evaluation studies provide evidence that, in Christian adults, person-based and embodied conceptions of God coexist with theologically informed trans-personal and disembodied conceptions (Barlev et al., 2018). The naïve conception is not always suppressed in favour of the more erudite conception, however. Rather, people switch between explanations depending on the context (Lane et al., 2016; Legare & Gelman, 2008; Watson-Jones et al., 2017). For example, people give a supernatural account of death in a religious setting but a biological account in a secular setting (Harris & Gimenez, 2005). Representational co-existence can thus lead to 'explanatory co-existence': People may recruit different conceptions to give different explanations of the same phenomenon, in different contexts.

'Fragmentation accounts of belief storage' (Kindermann & Onofri, 2021; Leiser, 2001; Sommer et al., 2023) elaborate the notion of representational co-existence. On these accounts (e.g., Bendaña & Mandelbaum, 2021), belief representations are stored in a large number of distinct data structures ('fragments') that are causally isolated: they are independently formed (through different processes or in different contexts), independently accessed (typically activated by different stimuli, in different contexts), and independently updated in the light of new information (e.g., further scientific instruction only updates the representation of the scientific theory). Due to constraints on working memory (review: Oberauer et al., 2016), these belief sets are checked, at most, for internal consistency. As a result, people may hold different, mutually incompatible conceptions of the same phenomenon (say, a naïve and a scientific theory, stored in different 'fragments').

Vision is one of the phenomena about which people hold conflicting theories, which may be co-represented in distinct belief fragments. According to *intromissionist* conceptions of vision, vision involves light rays entering our eyes. According to *extramissionist* conceptions, vision involves force rays leaving our eyes. The prevalence of these conceptions has been studied through tasks including agreement rating and drawing tasks (review: Winer et al., 2002), and with sophisticated behavioural measures (Guterstam et al., 2019). These studies found that both conceptions are maintained by laypeople, sometimes by the same individual.

Extramissionist and intromissionist beliefs about vision are typically taken to represent a psychological conflict between naïve and scientific theories of vision. Extramissionism has the hallmarks of *intuitive* naïve theories: It has been maintained across different historical periods (Lindberg, 1976), and is today found across different age groups (Piaget, 1979; Winer et al., 1996) and cultures (Dundes, 1981). Intuitive naïve theories persist not merely in the face of scientific counter-instruction, but also in the face of neural degeneration: In speeded verification studies, the response patterns of Alzheimer patients revert to those of pre-schoolers (Lombrozo et al., 2007; Zaitchik & Solomon, 2008). The natural thought is that intuitive naïve theories are formed due to largely innate cognitive propensities, underwritten by largely automatic cognitive processes, that may, e.g., create cognitive or perceptual illusions (McCloskey et al.,

1983), and therefore strike people as intuitive, across different cultural and historical settings, and persist where learning is lost.

Intuitive naïve theories have been repeatedly articulated by speculative philosophical accounts formulated prior to scientific inquiry (McCloskey, 1983). Thus, extramissionism has been articulated by ancient philosophers including Empedocles, Plato, and Euclid (see Gross, 1999; Meyering, 1989). Motivated by a rather different rationale (set out above), we now turn to more recent philosophical debates, to identify a belief conflict that fits less easily into the mould of the naïve vs scientific theory conflicts that have been examined by extant studies.

1.2 Philosophical context: indirect realist theories of perception

Philosophical debates about aporetic problems concerning the nature of perception have been shaped by a contest between Direct Realism and Indirect Realism (reviews: Crane & French, 2021; Pautz, 2021; Robinson, 1994). The central claim of *Indirect Realism (IR)* is that people see physical objects, like tables or tomatoes, *by* or *in virtue of* being aware of something distinct from them, variously called mental images, 'ideas', or 'sense-data'. At least typically, Indirect Realists argue that when our visual experiences are *veridical* (i.e., when things look their true size, shape, and colour to the observer), the physical object of sight causally brings about this mental image (Russell, 1912/2000; Jackson, 1977; Robinson, 1994). The main strand of philosophical IR, often traced back to Locke's synthesis of early modern science (Locke, 1700/1975), builds on this metaphysical claim:

> *Perceptual Indirect:* People see physical objects in their environment in virtue of being aware of images (ideas, sense-data) in their mind; namely, they are aware of images that are caused in their mind by their interaction with physical objects.[3]

Often, this metaphysical claim is accompanied by one or two epistemological claims. According to the first, the acquisition of information through vision involves inference:

> *Phenomenal Indirect:* Viewers infer the existence and nature of any physical object of sight from the occurrence and nature of the images (ideas, sense-data) the object causes in their minds.

For example, when you 'see' a book lying on a table, you experience a mental image of a book lying on a table, and infer the presence of a book on the table as the cause of the experience. Philosophical theories in line with IR often present this inference

[3] While the ambiguous phrase 'to be aware' is common in metaphysical statements of IR, it is usually *not* used in a strong epistemic sense that would imply that viewers know that they are seeing mental images. As we argue below, the philosophically intended claim seems to be rather similar to the claim that vision involves seeing mental images, which is part of the folk IR theory we will assess.

as automatic and unconscious (e.g. Berkeley, 1709/1975; Russell, 1912/2000; Price, 1932).

The view that inference is required for visually acquiring information about the nature and existence of objects in the environment often goes together with the view that our perceptual experiences themselves only inform viewers how things 'appear' or 'look' to them; i.e., these experiences, by themselves, do not provide people with knowledge, or justified true beliefs, about how things actually are.

> *Epistemic Indirect:* On the basis of their perceptual experience, viewers are justified in holding beliefs about how things appear to them, and not about how things are.

According to this claim, if it looks to us as if there is a lemon, this experience does not, by itself, justify believing that there is a lemon, because things would look exactly the same to us, if we looked at, say, a lemon-shaped piece of soap. On this view, 'appearances' provide at best 'evidence' for the nature and existence of things in our environment (Ayer, 1956/1990; and for discussion; Austin, 1962).

Direct Realism (DR) amounts to the rejection of the three claims just canvassed. Traditionally, it has been regarded as reflective of the 'common-sense' conception of vision (e.g., Allen, 2020; Martin, 2002; Searle, 2015; Siegel, 2011), which philosophers, often implicitly, equate with the reflectively endorsed folk beliefs of scientifically untrained adults (e.g., Snowdon, 1981, p. 176), and often take to be widely, or even universally, shared (e.g., Lewis, 1997, p. 325; Strawson, 1959, p. 10). This is true of both philosophers who accept some form of DR (e.g. Reid, 1785/1969; Austin, 1962; Martin, 2002; Searle, 2015) and those who reject it (e.g., Hume, 1739-40/2011; Russell, 1912/2000; Boghossian & Velleman, 1989).

DR so understood is neutral between contemporary naïve realist and representationalist theories of perception. According to naïve realists, perceptual experiences are fundamentally relational, constituted by a relation of conscious acquaintance between perceivers and physical objects and their properties (e.g., Allen, 2020; Fish, 2009; Genone, 2016; Martin, 2002). According to representationalists, by contrast, perceptual experiences are content-bearing states or events that represent physical objects and their properties (e.g., Pautz, 2021; Searle, 2015; Siegel, 2011). Representationalism is typically considered to be a 'direct realist' theory in the sense that we are not aware of physical objects and their properties in virtue of being aware of anything else, such as mental images or sense-data (e.g., Crane & French, 2021; McLaughlin, 2010). Which of these philosophical theories better captures 'the' common-sense conception of vision is a matter of debate (e.g., Allen, 2020; Martin, 2002; Raineri, 2021).

While it is generally assumed that, at any rate prior to philosophical elaboration, DR is common sense, the status of IR is controversial. IR is typically assumed to be in need of philosophical or scientific motivation: for example, by arguments from illusion or hallucination (e.g., Hume, 1739-40/2011; Robinson, 1994), reflection on scientific theories of sensory qualities like colours (e.g., Boghossian & Velleman, 1989; Jackson, 1977), or broader reflections on scientific theories of vision (e.g., Hoffman et al., 2015; Hohwy, 2013). However, there is disagreement about whether mod-

ern scientific theories of vision are actually best understood as forms of IR: Against popular-scientific interpretations of these theories in line with IR (e.g., Hawking & Mlodinov, 2010, pp. 62–63), some philosophers argue that these scientific theories are representationalist forms of DR (e.g., Burge, 2005; Drayson, 2021), while others argue that they support neither philosophical theory (McDowell, 2010). The latter assessment is consistent with empirical findings that laypeople regard neither DR nor IR as scientifically accurate (Fischer et al., 2026).

We suggest that, like DR, also IR is reflective of pre-scientific folk beliefs. The pertinent beliefs are particularly evident from early modern statements of IR, as in this account from Locke (1700/1975, II.iii.1):

> Thus light and colours … come in only by the eyes: All kinds of noises … only by the ears: The several tastes and smells, by the nose and palate. And if the organs, or the nerves, which are the conduits to convey them from without [i.e., from outside] to their audience in the brain, the mind's presence room [literally: monarch's reception room] (as I may so call it) are any of them so disordered as to not perform their functions, they have no postern [i.e., no small rear gate] to be admitted by; no other ways to bring themselves into view, and be perceived by the understanding.

As Daniel Dennett (1991, p. 107) comments, 'Many theorists would insist that they have explicitly rejected such an obviously bad idea. But … the persuasive imagery … keeps coming back to haunt us—laypeople and scientists alike'. Thus, Locke explains that talk of spaces ('room') and seeing ('bring into view') in this context is not to be understood literally, but frequently makes inferences that are only supported by literal interpretations (Fischer, 2011, pp. 35–40, 53–61, 103–108).

We infer that an intuitive naive account of vision can be recovered through literal interpretation of this supposedly scientifically informed philosophical account. Literal interpretation yields a *'Cartesian Theatre conception'* (Dennett, 1991) consisting of two components: (1) The notion of a *sensorium commune* (articulated already by Aristotle), where input from the five senses is integrated into one unified perception, and (2) the idea that this unified perception is *seen*, etc., with an *inner eye*, etc., in an *inner theatre*. Studies using drawing tasks and verbal agreement ratings have found evidence of widespread acceptance of this way of thinking: Laypeople locate consciousness (but not unrelated neurological processes or unconscious thinking) in a single, confined area in the prefrontal cortex (behind the eyes), and they take conscious experience to occur only after (and arguably, as a result of) complete neural processing of sensory input (Forstmann & Burgmer, 2022; cf.; Bertossa et al., 2008). One study probed for beliefs about where and when 'the brain creates the subjective, conscious experience of seeing a tree' by combining verbal statements with illustrations (tree with blurry edges) that suggested this experience involved seeing a mental image (Forstmann & Burgmer, 2022, Study 4). Another study used agreement rating tasks with verbal and pictorial stimuli to probe this key idea and found widespread acceptance of the folk IR idea that vision involves 'seeing' mental images (40–50% of a lay sample agreeing with verbal statements of this idea) in a kind of inner theatre

(with 70% agreeing that illustrations of the Cartesian Theatre accurately describe what happens in vision) (Fischer et al., 2023).

We therefore suggest that a *folk Indirect Realism* about vision can be recovered from the key metaphysical and epistemological claims associated with philosophical IR (Perceptual Indirect, Phenomenal Indirect, and Epistemic Indirect) through concrete or literal interpretation of their key terms 'awareness', 'inference', and 'idea' (whose literal meaning, in early modern English, the *Oxford English Dictionary* gives as 'image'). On such an interpretation, these terms refer to seeing mental images and making conscious mental operations.

Against the philosophical orthodoxy that 'the' folk conception of vision is Direct Realist, we hypothesize that:

H_1 Laypeople are collectively torn between this folk Indirect Realism about vision and a conflicting Direct Realist conception.

H_2 Many laypeople are individually torn between these two conceptions.

Either or both of these opposing folk conceptions may turn out to have all the attributes of intuitive naïve theories (see Sect. 1.1). For now, however, we remain agnostic about their precise status (which we will discuss in Sect. 3), and think of these conceptions simply as internally coherent sets of beliefs.

H_1–H_2 address our second research question (RQ2). The next two hypotheses **H_3** and **H*** address the more fundamental (RQ1): Provided that we are dealing with beliefs (as per **H*** below), we suggest inter- and intrapersonal conflicts are due to belief fragmentation (Sect. 1.1), and hypothesize:

H_3 Laypeople can draw on two internally coherent conceptions of vision that are stored in different belief fragments.

1.3 Validating a belief inventory

Hypotheses about folk beliefs are typically assessed with belief inventories which contain verbal statements of the beliefs of interest and agreement ratings. To assess our hypotheses, we will develop and validate a new *Direct/Indirect Realist Belief Inventory (DIRBI)* that employs agreement rating tasks with verbal and visual stimuli. To 'validate' a psychometric measure is to justify the intended interpretation of the responses that have been elicited by the measure from the samples used; this is typically done through abductive reasoning that shows that the observed responses are better explained by the intended interpretation than by available competing explanations (Messick, 1995; Zumbo, 2023). Before we are able to use our new measure to assess **H_1**–**H_3**, we therefore need to examine the hypothesis

H* The observed responses to our inventory are due to, and accurately reflect, respondents' beliefs about, or conceptions of, vision.

Efforts to validate philosophically relevant belief inventories have been most extensive in connection with beliefs about free will, where validation efforts were mainly driven by concerns about theory contamination and focused on assessing item formulation and selection (review: Nadelhoffer et al., 2014). The more fundamental concerns we canvassed at the outset – whether high agreement ratings for two competing conceptions of the same phenomenon reflect genuine belief – motivate us to execute standard validation protocols more comprehensively, and to start at the beginning, by explicitly defining the relevant construct of 'belief' (an essential task frequently overlooked in the development of belief inventories). This will motivate empirical investigation of the inferential deployment of the posited beliefs (a previously entirely overlooked task).

In cognitive science, beliefs are typically conceptualised as propositional attitudes: *beliefs* are mental states bearing an informational content ('proposition'), and to have a belief is to have a certain attitude towards this content, viz., to regard it as true (Albarracin & Pitliya, 2022; Porot & Mandelbaum, 2021; Quilty-Dunn & Mandelbaum, 2018). Most psychological research on naïve theories (Sect. 1.1) assumes a *representationalist* construal of these propositional attitudes. So does most philosophy of cognitive science (since Dretske, 1988; Field, 1978; Fodor, 1987). On this construal, to believe some content (e.g., the proposition that people see physical objects in virtue of being aware of mental images) is to have a representation with this content stored in the brain, available for deployment in cognitive tasks including reasoning, problem solving, decision making, and action planning. Without theories identifying conditions under which the available information is deployed, representationalism does not provide belief attributions with behavioural implications. Apparently mitigating this shortcoming, representationalism is traditionally combined with a functionalist perspective, which seeks to characterise beliefs by their causal relations to environmental stimuli, other mental states, and behaviour. However, these causal relations have remained underspecified (Schwitzgebel, 2023; §1.4), and more recent research on the functions of belief has identified very different functions (Funkhouser, 2017; Meyer & Knobe, 2026; Sommer et al., 2023). It therefore does not seem profitable to treat 'belief' as standing for states which all share one kind of causal role.

We accordingly take a cautious bottom-up approach and start out with a construct that is mainly constrained by the rationale informing our hypotheses. We have suggested that speculative philosophical theories develop the beliefs of interest. This implies that these beliefs are *accessible*, so that agreement rating tasks will tap into them. We further take these beliefs to have the two properties most frequently regarded as *constitutive* of belief (Quilty-Dunn & Mandelbaum, 2018): Firstly, these beliefs are *standing* or 'dispositional', rather than occurrent states; they persist even when their content is not deployed in any cognitive task. Secondly, their content *can* be *deployed in judgment and inference*. While it remains to be determined in which everyday situations the beliefs of interest are deployed, IR beliefs have been deployed by philosophers, in writing and in introductory classes, to set up intuitively compelling philosophical problems, including sceptical problems (e.g., Descartes, 1641/1996; Ayer, 1956/1990; Stroud, 1984). We infer that freshmen and other lay-

people will deploy these beliefs in reasoning at least when faced with topically relevant philosophical tasks.[4]

On the common representationalist perspective, we need to distinguish between 'core beliefs' which are explicitly stored and represented, and 'tacit beliefs' which are automatically inferred from them on demand (Crimmins, 1992; Frankish, 2004). To ensure intelligibility and naturalness of our scale items, we will not use abstract terms like 'physical objects' to phrase the statements of interest, but will rather provide for each hypothesised abstract belief (like *Perceptual Indirect* or *Phenomenal Indirect*, above) several more concrete statements about what happens when someone sees a banana, a table, or another prototypical physical object of sight (Sect. 2.2.1).[5] Participants are unlikely to have explicitly considered previously the question of what happens when someone sees a banana, or to have stored a belief about this specific matter. Within a representationalist framework, our hypothesis $\mathbf{H}^*$ is, more specifically, that agreement ratings for concrete items are expressive of tacit beliefs which are automatically derived from the core beliefs captured by the more abstract philosophical statements.

2 Three studies

We conducted three questionnaire-based studies to assess $\mathbf{H}^*$. Studies 1–2 simultaneously served to assess $\mathbf{H_1}$–$\mathbf{H_3}$.

2.1 Overview – Studies 1–3: approach and predictions

Studies 1–3 served to develop and validate the *Direct/Indirect Realist Belief Inventory (DIRBI)*. This inventory is intended to capture potentially conflicting beliefs between which laypeople may be torn both collectively (as per $\mathbf{H_1}$) and also individually (as per $\mathbf{H_2}$). Unlike most psychometric measures, it is therefore not intended to place individuals on a continuum – say, how strongly 'Direct Realist' or 'Indirect Realist' they are. Unlike many recently proposed scales, the DIRBI also ventures into an underexplored domain, where no measures for related constructs exist already. We accordingly modified a standard three-stage psychometric validation protocol, proposed by Loevinger (1957) and developed by Simms (2008).

The first or *'substantive validity'* stage serves to define the construct to be measured and to develop an initial item pool. In addition to defining the relevant construct of 'belief' (Sect. 1.3), this requires identifying topically relevant claims that could plausibly be part of a DR or IR conception of vision and could make the relevant difference between the two conceptions. To identify such beliefs, two authors

[4] By contrast, nascent process models of belief suggest that philosophically popular questions about the evidence-sensitivity and in/voluntary nature of beliefs may not be well-framed, as belief appraisal in acquisition and updating may involve several different cognitive processes, some of which may possess, and some may lack, the property of interest, which can therefore not be attributed to beliefs *tout court* (review: Soter, 2025). We therefore do not include these further properties in our construct.

[5] Interestingly, Fischer et al. (2026) found that lay participants do not deem such concrete statements less scientifically accurate than more abstract statements that speak of seeing a 'physical object'.

with teaching and research experience in philosophy of perception examined recent and influential literature reviews and textbook treatments of the key debates about the nature and epistemology of perception that pit DR against IR (Crane & French, 2021; Fish, 2009; Lyons, 2023; Pautz, 2021; Robinson, 1994). We took stock of the direct/indirect distinctions drawn in this corpus. In addition to the three distinctions of Perceptual, Phenomenal, and Epistemic Indirect vs Direct explained above (Sect. 1.2), these included, e.g., distinctions between Referential In/Directness and In/Direct World Involvement (Lyons, 2023). We eschewed distinctions that relied on substantive philosophical theorising (like the last two mentioned). We also eschewed a belief (the 'Phenomenal Principle', Robinson, 1994) whose acceptance and rejection, respectively, are associated in the literature with IR and DR, respectively, but which Fischer et al. (2023) had found to load on a different factor than the key metaphysical tenants of these conceptions. We discussed the remaining DR vs IR notions informally in undergraduate classes to gauge their intelligibility. Further material development was supported by pre-structured interviews with lay informants (below, Sect. 2.2.1).

The second or *'structural validity'* stage serves the psychometric evaluation of the items and the construction of the scale. The psychometric evaluation initially focuses on examining the factor structure of the nascent scale: to what extent ratings for different items are influenced by the same construct (e.g., by the same belief or conception), and the internal consistency of the scale and its constituent subscales, i.e., to what extent the ratings for different items included in them co-vary. The items designed to articulate different tacit beliefs inferred from the same core belief should load on the same factor and show high internal consistency. Given the narrowness of our construct (a single 'core belief' per subscale), we aim for subscales whose items load on one factor only, and do so ≥ 0.80, and with Cronbach's $\alpha \geq 0.85$, even where subscales have few items (which automatically reduces α). Psychometric evaluation further assesses to what extent the factor structure remains stable across different samples, suggesting that when different people take the survey, on different occasions, the same constructs influence their responses (e.g., the same beliefs), as per **H***. Finally, it assesses the test-retest reliability, i.e., to what extent the same participants give similar responses on different occasions. The *standing* – and *accessible* – nature of the belief states posited by $\mathbf{H}^*$ predicts high test-retest reliability. Studies 1–2 will examine the factor structure of the nascent DIRBI. Studies 2–3 will examine test-retest reliability.

The third or *'external validity'* stage typically involves comparisons with different measures: To provide further evidence that the proposed scale measures the intended construct, developers seek to show that it yields similar results as measures of similar constructs ('convergent validity'), different results than measures of different constructs ('discriminant validity'), and the same results as an established benchmark measure of the construct ('criterion-related validity'). This is not feasible where there are as yet no measures of similar constructs, let alone an established 'benchmark' criterion. It is also unfeasible where links to potentially related constructs (e.g., to other beliefs about vision) are too poorly understood to support predictions.

This is the case here. Cognitive neuroscience and philosophical history motivate conflicting predictions about the relationship between DR and IR on the one hand,

and extra- and intromissionism, on the other. Attention schema theory suggests that extramissionism is rooted in an implicit schema of attention (Graziano, 2022; Guterstam et al., 2019). This also affords a plausible explanation of IR beliefs (Sect. 3), and suggests that IR and extramissionism should go together. But all prominent IR theories in philosophy have integrated their ideas with intromissionism (e.g., Russell, 1912/2000; Robinson, 1994), and some DR theorists have regarded extramissionist beliefs as reflecting the phenomenological character of experience (e.g., Allen, 2019). Thus, the relationship between DR/IR beliefs and extra-/intromissionist beliefs remains unclear.

Instead of examining the relationship to other constructs and criteria, Study 3 will therefore complete initial validation by assessing the remaining key feature of the belief construct we use in $\mathbf{H}^*$: It will examine whether the putative beliefs measured by our inventory are deployed in judgment and reasoning tasks that are philosophically relevant.

Studies 1–2 will simultaneously address $\mathbf{H_1}$-$\mathbf{H_3}$. Philosophers practically unanimously regard DR as 'the' common sense view of vision (Sect. 1.2). Against this orthodoxy, $\mathbf{H_1}$ maintains that laypeople are collectively torn between a DR and an IR conception of vision. This predicts that

[Prediction 1] A higher proportion of lay participants than can be plausibly explained by performance error will endorse DR and IR, respectively.

$\mathbf{H_2}$ claims that many laypeople are individually torn between those conceptions. This motivates

[Prediction 2] A non-negligible proportion of laypeople will endorse both DR and IR conceptions simultaneously.

(This will happen even if some torn souls may effortfully impose consistency on their responses.) $\mathbf{H_3}$ suggests the conflicting conceptions are stored in different belief fragments. This predicts a certain factor structure: If ratings for conflicting DR and IR items are driven by the same set of beliefs (say, widely, but not universally, shared DR beliefs that lead participants to agree with Perceptual Direct items and to disagree with Perceptual Indirect items), then ratings for opposing items should load on the same factor (positively vs negatively). By contrast, if (as per $\mathbf{H_3}$) these ratings are driven by different beliefs, stored in different fragments,

[Prediction 3] Ratings for opposing DR and IR items should positively load on different factors.

Finally, fragmentation accounts of belief storage also suggest that different stimuli or stimulus formats can activate different fragments (Bendaña & Mandelbaum, 2021). We therefore ask whether a non-negligible proportion of participants will endorse a conception of vision when presented by an illustration, but not a verbal statement (cf. Winer et al., 1996).

We now report three studies with analyses that address **H*** and support the development of our new DIRBI measure (Sects. 2.2–2.4). We will then report further analyses, conducted on data from Studies 1–2, that examine predictions from hypotheses $\mathbf{H_1}$–$\mathbf{H_3}$ (in Sect. 2.5).

2.2 Study 1

2.2.1 Study 1 – Methods

Participants ($N = 500$) were recruited via Prolific and remunerated. They were between 18 and 78 years old (mean 29 years). 179 were male, 309 female, 9 non-binary, and 3 'preferred not to say'. All were resident in the UK. Screening participants by education level and first degree, we recruited participants who had either no university education or university education in disciplines other than natural sciences, psychology or philosophy. Even so, half (49%) of participants also had higher-level education (secondary school 'A-level' or higher) in these disciplines. Almost all participants had at least some university education.

Materials and procedure: We administered an agreement rating task online, through *Qualtrics*. Participants were presented first with verbal items and then with animations. Material development observed the basic item writing principles codified by Simms (2008, Table 1) and was supported by a literature review (see above) and semi-structured interviews.[6]

Verbal items. We used 48 verbal items in 8 conditions or subscales. Each condition or subscale had 6 items that referred to different objects of sight but were intended to capture the same metaphysical or epistemological 'core belief' about vision. Four conditions probed a core DR belief and four corresponding conditions probed the opposing IR belief:

Table 1 Study 1 – exploratory factor analysis: factor loadings on the eight verbal item categories. Significant loadings (≥0.22 for samples of 500; Stevens, 2012) highlighted in bold

	Factor 1	Factor 2
Perceptual Direct (PD)	**0.909**	−0.201
Perceptual Indirect (PI)	**−0.412**	**0.792**
Perceptual Causal Direct (PCD)	**908**	−0.216
Perceptual Causal Indirect (PCI)	**−0.392**	**0.796**
Phenomenal Direct (PHD)	**0.787**	−0.128
Phenomenal Indirect (PHI)	**−0.250**	**0.714**
Epistemic Direct (ED)	**0.665**	−0.216
Epistemic Indirect (EI)	0.073	**0.834**
% variance explained	53.3	18.2
Eigenvalue	4.26	1.46

[6] Twelve UK informants were asked (i) to indicate on a 5-point scale how easy it would be to recognise various objects by looking, (ii) to rate the intelligibility of a representative sample of draft verbal items, (iii) to draw what happens when someone looks at an apple, and (iv) to interpret images representing what happens when we see an apple. Responses guided the development of verbal items and the design of animations.

Perceptual Direct [PD]: e.g., 'When you look at an orange, you see just the orange and not a mental image of the orange.'

Perceptual Indirect [PI]: e.g., 'When you look at a tomato, you see a mental image of the tomato and not just the tomato.'

Perceptual Causal Direct [PCD]: e.g., 'When you look at a cucumber, you see just the cucumber and not a mental image that is caused by the cucumber.'

Perceptual Causal Indirect [PCI]: e.g., 'When you look at a strawberry, you see a mental image that is caused by the strawberry and do not see just the strawberry.'

Phenomenal Direct [PHD]: e.g., 'When you look at an apple, you just see there is an apple and do not consciously work out that it is an apple.'

Phenomenal Indirect [PHI]: e.g., 'When you look at a lemon, you consciously work out that it is a lemon and do not just see there is a lemon.'

Epistemic Direct [ED]: e.g., 'When you see a potato, you know that it is a potato and not just that it looks like a potato.'

Epistemic Indirect [EI]: e.g., 'When you see a banana, you know just that it looks like a banana and not that it is a banana.'

While items within the same condition differed in the object of sight referred to, all items referred to medium-sized objects with stereotypical shapes and colours, that interviewees in the semi-structured interviews had deemed easily recognizable by looking (as revealed by mean agreement ratings$\geq$4.5, on a 5-point scale). Within each item category, there was one fruit or vegetable, one piece of furniture, one tool, one animal, one vehicle, and one thing you might find on a table. A list of verbal items is provided in Appendix A.

Verbal items were presented in random order, and participants rated their agreement with these statements on a 7-point scale. To guard against acquiescence bias (Jackman, 1973), two versions of our questionnaire presented response options in different order: For half the participants, vertically ordered response options were presented with 'Strongly Disagree' at the top and 'Strongly Agree' at the bottom; for the other half, response options were reversed.

Animations. Following the verbal items, participants were presented with 7 animations, intended to represent DR and IR views of what happens when we perceive an apple. Animations were co-produced with artist Alexandra Carr. The design took into account information garnered through the norming study, including participants' drawings, explanations, and feedback on draft images. All animations started with an outline of a head and eye appearing on the right-hand side of the screen. After this, an apple appeared on the left-hand side of the screen. In the first four animations (retained for Studies 2–3), an arrow moved from the apple to the eye, and this was

followed either by cogs turning inside the head ('Cogs'), a screen with an apple on it appearing inside the head ('Screen'), a screen *and* subsequently cogs appearing inside the head ('Cartesian'), or nothing further happening when the arrow reached the eye ('Direct'). One animation featured an arrow moving from the eye to the apple ('Arrow Out'). Two further animations used an extending spatter pattern, moving from the apple to the eye or from the eye to the apple. As an example, Fig. 1 provides a still from the 'Cartesian' animation. Stills from, and links to, all animations are provided in Appendix B.

In the main study, stills of all 7 animations were first presented simultaneously on the same screen. After that, animations were presented individually, in a fixed order intended to make salient the relevance of omissions: Cogs, Screen, Direct, Cartesian, Arrow Out, Spatter Out, Spatter In. For each animation, we asked two questions: First, we probed participants' interpretation by asking, 'Which, if any, of the following statements describe the animation? Tick as many as apply.' Response options included statements corresponding to Perceptual Causal Direct, Perceptual Causal Indirect, Phenomenal Direct, and Phenomenal Indirect, as well as the option 'None of the above/I do not understand the animation'. Second, we asked participants to rate on a 7-point scale 'how accurately this animation represents what happens when we see an apple'. Finally, participants were shown stills of all 7 animations, asked to indicate which one 'best represents what happens when we see an apple', and to explain their choice in a free-text comment.

Compatibility. We finally wanted to find out whether participants perceive DR and IR items as mutually incompatible (as philosophers tend to do), or, e.g., take them to employ the verb 'see' in different senses, thus avoiding conflict. Once participants

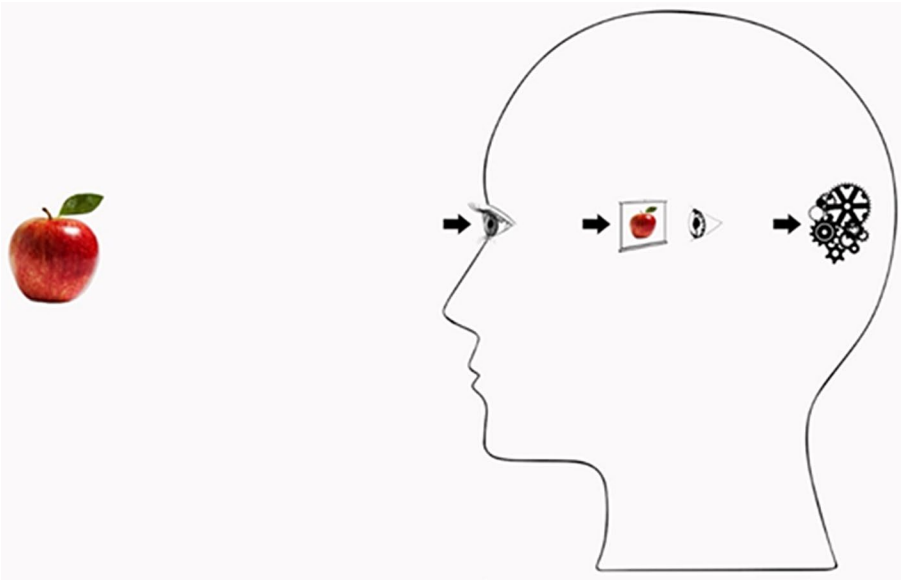

Fig. 1 Still from 'Cartesian' animation used in Studies 1–3

had completed the above tasks, they were therefore presented with 6 pairs of sentences and asked:

> Are the sentences in each pair compatible with each other? In other words: Can both of them be true together? Or must one be false, if the other is true? We are not interested in whether you think any of them is actually true. Rather, we want to know how you think the two sentences are related: Are they compatible with each other? Or are they incompatible each other?

Three of these sentences expressed opposing DR and IR views about the metaphysics of perception (Perceptual Direct/Indirect), whether perception involves inference (Phenomenal Direct/Indirect), and whether objects or appearances are epistemically basic (Epistemic Direct/Indirect), for example:

> [*Perceptual*] When you look at a bus, you see just the bus and not a mental image of the bus.
> When you look at a bus, you see a mental image of the bus and not just the bus.

The other three pairs of sentences employed perceptual verbs in different senses, e.g.:

> When in a spiritual crisis, people are very welcome to come and see what the teachings are about, without any obligation.
>
> When in a spiritual crisis, it is important to see the people that matter in your life and talk to them.

These items were taken from Fischer et al. (2023, Appendix G), who had based them on usage examples from the *British National Corpus* and had found that participants from a similar sample deemed them compatible. After reading the above instructions, participants were asked for each pair, 'Could the following two sentences both be true together?', and responded on a 7-point scale (strongly disagree to strongly agree).

Analysis: We scrutinized participants' responses for evidence of non-engagement (e.g., sustained series of identical responses to different items) but found none. We also examined responses to each subscale for order effects arising from the different ordering of response options. Independent samples t-tests revealed no significant effects, with p-values ranging from $p=0.17$ for PD to $p=0.92$ for PI. We also examined responses to each subscale for differences between participants with and without higher-level education (A-level or above) in natural sciences, psychology, or philosophy. Independent samples t-tests revealed a significant difference for PCI ($t(498)=2.196$, $p=0.03$) and a marginal difference for PI ($t(498)=1.676$, $p=0.09$), where the scientifically educated provided higher mean ratings. However, the effect sizes remained negligible (d's<0.2), and no significant effects were observed in the other conditions (all p's≥ 0.26). We therefore analysed data from all participants together.

We scrutinized verbal items through Principal Component Analysis, and assessed the internal consistency of each condition or subscale through Cronbach's alpha. No outliers emerged. Findings also supported item development for Study 2.

To further assess $\mathbf{H}^*$, we examined the factor structure of the observed agreement ratings. To do so, we conducted an exploratory factor analysis (EFA), and then tested initial hypotheses through confirmatory factor analysis (CFA) using structural equation modelling.

2.2.2 Study 1 – Results: factor analyses

We first analysed responses to **verbal items**. (Analyses of responses to animations are reported in Sect. 2.5.)

Principal Component Analysis identified only one factor with Eigenvalue ≥ 1 per subscale, but in most subscales some items failed to attain the sought loadings ≥ 0.80, while all subscales reached the sought level of internal consistency (Cronbach's alpha ≥ 0.85). We infer that the items of each subscale tap into one and the same underlying belief structure (e.g., *When someone looks at a physical object, they see just the object, and not a mental image of it*) but that ratings for several items were still unduly influenced by other factors. For details, see Appendix A (Table A1).

25 'extreme responders' used only response options '1' and '7', throughout. This resulted in a skewed distribution. We therefore used the recommended transformation for non-normal variables (Tabachnick & Fidell, 2019) for all factor analyses, though this did not fully correct for non-normality.

Exploratory factor analysis (using principal components extraction and Varimax rotation with Kaiser normalisation) revealed two factors with Eigenvalues over 1. These were retained in line with the Scree test (Cattell, 1966) (see Appendix C). The two factors explained approximately 71.5% of the variance in verbal item ratings (see Table 1)—well above the 60% threshold commonly employed in the social sciences (Hair et al., 2014). The significance of factor loadings depends upon the sample size; for samples of 500, loadings of 0.22 are deemed significant (Stevens, 2012). Each condition significantly loaded on one of the two factors. There were no cross-loadings. Four conditions loaded positively on each factor. The verbal part of the questionnaire thus meets common requirements for scale construction (Spector, 1992). Note that all DR items load positively on one factor (Factor 1), whereas all IR items load positively on the other (Factor 2). With one exception, IR items load negatively on Factor 1, whereas the negative loadings of DR items on Factor 2 remain shy of significance.

Confirmatory factor analysis used structural equation modelling to examine in the light of these findings whether all conditions or sub-scales are necessary to identify Direct Realists, Indirect Realists, and 'torn souls' (or whether, e.g., the conditions with the highest loading on each factor, like PD/PI or PCD/PCI suffice). While metaphysical and epistemological views aligned with DR and IR often go together in philosophy, they sometimes diverge. We therefore tested not only the 2-factor hypothesis suggested by our EFA but also the hypothesis that there are four underlying factors, dissociating the metaphysical and epistemological components of DR and IR, respectively. Strikingly, the 2-factor solution recognizing a DR and

an IR factor clearly failed to achieve an acceptable fit ($\chi^2 = 543.16$, df = 19, $p = 000$; CFI = 0.83; RMSEA = 0.235), while the 4-factor solution that dissociates the metaphysical and epistemological components had still poor fit ($\chi^2 = 199.85$, df = 14, $p = 0.000$) but met at least relative fit criteria (CFI = 0.94) though not absolute fit criteria (RMSEA = 0.163). For details, see Appendix C. To foreshadow, a model with excellent fit will be obtained from merged data from Studies 1–2 (Sect. 2.3.3).

2.3 Study 2

To develop a more concise measure of Direct/Indirect Realist beliefs about vision, and improve the structural validity of our measure (in particular, to ensure that no item ratings are unduly influenced by other factors than the beliefs of interest), Study 2 employed a more restricted set of verbal and image items to collect data from the same population.

2.3.1 Study 2 – Methods

Participants ($N = 319$) were recruited on the same basis as in Study 1. They were between 18 and 70 years old (mean 27.6 years, SD = 8.65). 110 were male, 201 female, 7 non-binary, and 1 'preferred not to say'. Almost all held BA or higher degrees. Half (50.6%) had higher education (A-level or above) in natural sciences, psychology, or philosophy.

Materials and procedure: We re-used verbal items and animations from Study 1. In the light of Study 1's inconclusive findings from structural equation modelling, we retained all 8 categories of verbal items. However, we deleted from each of the 8 verbal categories the two items with the lowest factor score, whose deletion also led to the greatest increase in internal consistency (Cronbach's alpha). This left 32 verbal items. We also omitted 3 of the animations from Study 1 which had attracted particularly high numbers of 'I don't understand the animation' responses in Study 1, and yielded the lowest proportion of intended interpretations. We retained 'Cogs', 'Screen', 'Cartesian' (Cogs and Screen), and 'Direct'. A list of verbal items and stills of, with links to, all animations are provided in Appendices A and B.

Participants were presented with the 32 verbal items in a random order. Participants used the same rating scale as in Study 1. After this, participants were presented with 4 animations in a fixed order: Cogs, Screen, Cartesian, Direct. The animations were otherwise presented in the same way as in Study 1, and with the same questions. To help us assess the test-retest reliability of our measures, 40 participants retook the same survey 4–5 weeks later.

Analysis: We conducted the same analyses as in Study 1. Again, we found no evidence of disengagement, and no order effect arising from the different ordering of response options, for any subscale: Independent samples t-tests revealed no significant order effects, with p-values ranging from $p = 0.31$ for EI to $p = 0.99$ for PHD. We also found no effect of higher education in natural sciences, psychology, or philosophy (all p's > 0.4).

Table 2 Study 2 – PCA factor analysis on each subscale (4 items each)

Subscale	Eigenvalue, Factor 1	% variance explained	Item loadings on Factor 1	Cronbach's α
Perceptual Direct (PD)	2.784	69.6	0.80–0.85	0.85
Perceptual Indirect (PI)	2. 976	74.4	0.84–0.88	0.89
Perceptual Causal Direct (PCD)	2.86	71.5	0.81–0.88	0.87
Perceptual Causal Indirect (PCI)	2.99	74.7	0.86–0.88	0.89
Phenomenal Direct (PHD)	2.71	67.9	0.79–0.85	0.84
Phenomenal Indirect (PHI)	2.83	70.7	0.84–0.85	0.86
Epistemic Direct (ED)	2.60	64.9	0.73–0.85	0.82
Epistemic Indirect (EI)	2.62	65.5	0.77–0.83	0.82

Table 3 Study 2 – exploratory factor analysis: factor loadings on the eight verbal item categories. Significant loadings (≥0.30, for samples of 300; Stevens, 2012) highlighted in bold

	Factor 1	Factor 2
Perceptual Direct (PD)	**0.933**	−0.094
Perceptual Indirect (PI)	**−0.616**	**0.595**
Perceptual Causal Direct (PCD)	**0.944**	−0.043
Perceptual Causal Indirect (PCI)	**−0.655**	**0.556**
Phenomenal Direct (PHD)	**0.622**	−0.229
Phenomenal Indirect (PHI)	−0.234	**0.771**
Epistemic Direct (ED)	**0.586**	−0.246
Epistemic Indirect (EI)	−0.024	**0.883**
% variance explained	52.9	16.1
Eigenvalue	4.230	1.284

2.3.2 Study 2 – Results: factor analyses

We analysed first responses to **verbal items.** (Again, analyses of responses to animations are reported in Sect. 2.5.)

Principal Component Analysis (PCA) identified only one factor with Eigenvalue≥1 per subscale. For all subscales except ED and EI, all items had loadings≥0.79 on this factor and Cronbach's α≥0.84. For all eight subscales, internal consistency was at least good (α≥0.82). For details, see Table 2.

Exploratory factor analysis (using principal components extraction and Varimax rotation with Kaiser normalisation, without prior transformations) revealed three factors with Eigenvalues over 1 (viz., 4.230, 1.284, and 1.085). Only two of these were retained, in line with the Scree test (see Appendix C). They jointly accounted for 69% of the observed variance. As in Study 1, each condition significantly loaded on one of the two factors, there were no cross-loadings, and four conditions loaded positively on each factor: all DR items loaded positively on Factor 1 and all IR items loaded

positively on Factor 2 (see Table 3). This time, however, the only negative loadings to attain significance were those of metaphysical IR items on the DR Factor 1.

Confirmatory factor analysis examined the same hypotheses as in Study 1, using structural equation modelling as before. As in Study 1, the 2-factor model failed to meet any fit criteria ($\chi^2 = 375.64$, df = 19, $p = 0.000$; CFI = 0.804; RSMEA = 0.243), while the 4-factor model displayed less poor fit ($\chi^2 = 126.60$, df = 14, $p = 0.000$) and met relative fit criteria (CFI = 0.94), but not absolute fit criteria (RSMEA = 0.16). For details, see Appendix C. While the factor structure of our questionnaire thus remains opaque at this point, it does seem stable across the two different samples of Studies 1–2, as required by **H***.

2.3.3 Follow-up factor analyses and discussion

To clarify the factor structure of our questionnaire, we sought to identify the sources of the discrepancies between exploratory and confirmatory factor analyses, and the poor fit of the structural equation models in Studies 1–2. To do so, we conducted further factor analyses. The stability of the factor structure and the similarity of the samples allowed us to combine the data from Studies 1 and 2 for these analyses.

We first performed an **exploratory factor analysis** using principal components extraction and Varimax rotation with Kaiser normalisation on data merged from Studies 1–2. Three factors had Eigenvalues over 1 (4.244; 1.374; 1.012). Two of these were retained in line with the Scree test (Appendix C). The two retained factors jointly accounted for 70.2% of the observed variance. Figure 2 visualizes how ratings from each subscale loaded on the two factors (for precise numeric values, see Appendix C, Table C1).

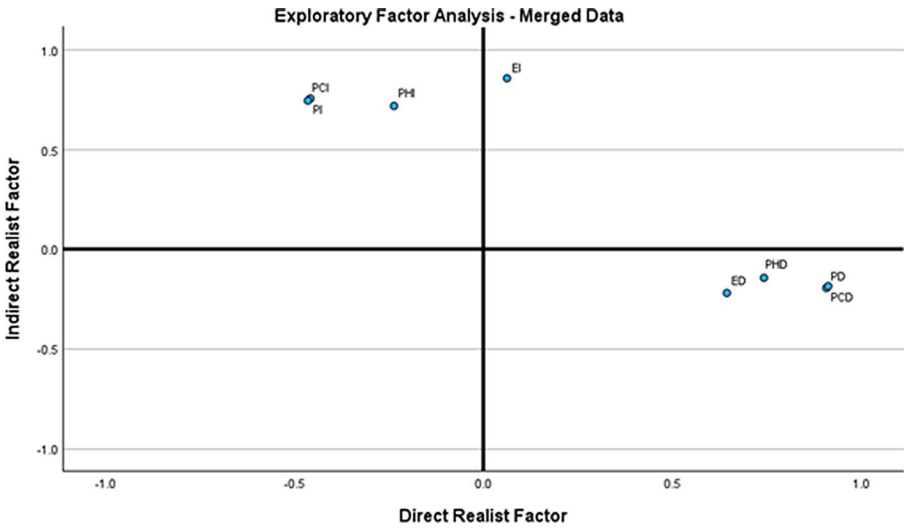

Fig. 2 Exploratory factor analysis on merged data from Studies 1–2 ($N = 819$): factor loadings of subscale ratings. X-axis shows loadings on Direct Realist factor. Y-axis shows loadings on Indirect Realist factor

We observe a clear separation between ratings for DR subscales (PD, PCD, PHD, and ED), on the one hand, and IR subscales (PI, PCI, PHI, EI), on the other. They fall into opposing quadrants in terms of their factor loadings (except for EI, which loads positively, but non-significantly, on the DR Factor). Loadings on both DR and IR factors are almost identical for the metaphysical subscales on each side (PD and PDC; PI and PCI). The Direct metaphysical subscales PD and PCD load a bit more strongly on the DR factor than epistemological subscales PHD and ED (see moderate divergence along the x-axis), while all load equally weakly and negatively on the IR Factor (see proximity on the y-axis). On the Indirect side, epistemological and metaphysical subscales load roughly equally strongly on the IR Factor (hardly diverge on the y-axis), but differ considerably in terms of strength of their mostly negative loadings on the DR Factor (considerably diverge on the x-axis). We infer these ratings are influenced to very different degrees by DR beliefs, as also borne out by the correlation table for ratings from the merged sample (Appendix C, Table C2).

These findings suggest that the poor absolute fit for the 4-factor models was due to the different extent to which DR beliefs influenced ratings for PHI and EI. To test this suggestion, we excluded those item ratings and sought to fit a 3-factor model to the remaining data (see Figure 3). This 3-factor model recognized a Metaphysical DR

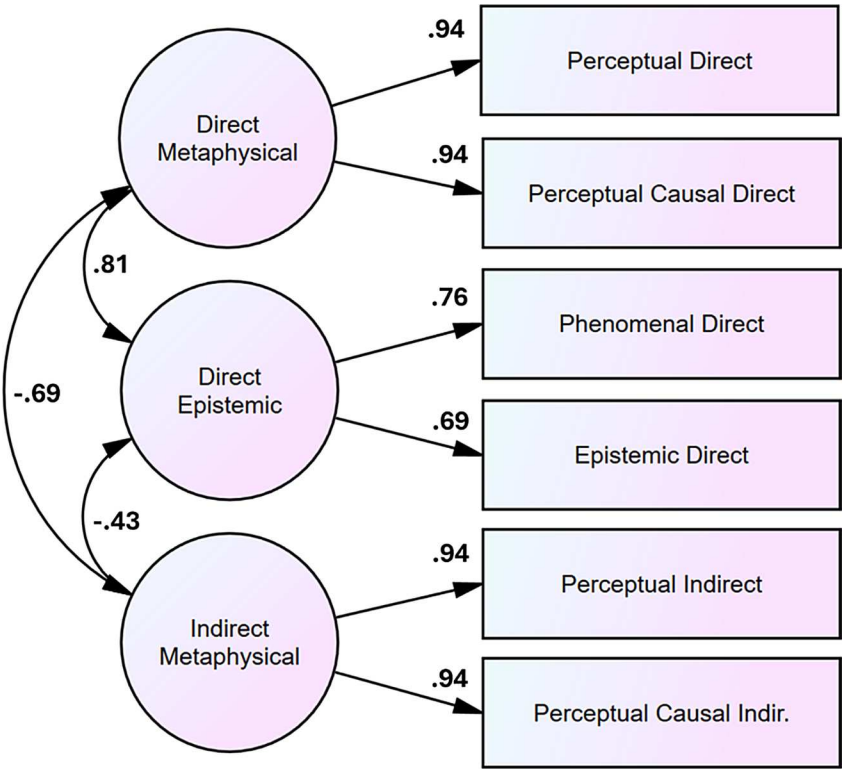

Fig. 3 3-Factor Structural Equation Model for merged data from Studies 1–2 (N=819), after exclusion of ratings from PHI and EI subscales, with correlations between factors shown on the left-hand side and factor loadings on the three factors specified in the centre

factor (on which PD and PCD both load with 0.94), an Epistemological DR factor (on which PHD and ED load with 0.76 and 0.69) and a Metaphysical IR factor (on which PI and PCI both load with 0.94). According to this model, the Metaphysical DR and Epistemological DR ratings are strongly correlated (0.81), while the – negative – correlations between Metaphysical IR, on the one hand, and Metaphysical DR and Epistemological DR, on the other, were weaker (−0.69 and −0.43, respectively). This model is visualised by Fig. 3. The fit of this model was excellent ($\chi^2 = 9.20$, df = 6, $p = 0.16$) and met both relative and absolute fit criteria (CFI = 0.999; RSMEA = 0.026).

The findings from Study 1, Study 2, and follow-up factor analyses are consistent with the hypothesis **H*** that agreement ratings for the items of these studies are largely due to beliefs about vision and, more specifically, that ratings for the items on each subscale express tacit beliefs that are inferred from the same core belief. Findings from Principal Component Analysis suggest that at any rate the four items retained for each subscale in Study 2 indeed tap into one and the same narrow construct: In six of the 8 subscales, all items had loadings ≥ 0.79 on Factor 1 (see Table 2). Only in the epistemic subscales did some items' factor loadings remain clearly shy (≥0.73) of the target value of 0.80, suggesting some ratings were non-negligibly influenced also by further factors. Despite the lower number of items in Study 2, internal consistency of all subscales was good ($\alpha \geq 0.82$). On the other hand, no subscale achieved values ≥ 0.90, which would suggest redundancy of items. We infer that none of our subscales should be shortened further.

The similarity of EFA and CFA results, respectively, in Studies 1 and 2, suggests that when different people take the survey, on different occasions, the same constructs influence their responses. This is consistent with **H*** and allowed us to merge data from the two studies. Looking at the merged data from Studies 1–2, we found very high correlations between PD and PCD and between PI and PCI (both 0.88). Accordingly, Confirmatory Factor Analysis conducted on this merged data observed very high loadings of PD and PCD as well as of PI and PCI, respectively, on the same factor, in a Structural Equation Model with excellent fit. This suggests that ratings for these items are due to the same constructs, for example, a DR and an IR metaphysical belief, respectively. For inventory construction, this implies that two of the four metaphysical subscales are surplus to requirements.

2.4 Study 3

Study 3 completed the initial validation of the DIRBI, i.e., the first assessment of the hypothesis **H*** that agreement ratings elicited by our inventory tap into genuine beliefs, understood as *standing* information-bearing states whose content is *inferentially deployed*.

2.4.1 Study 3 – Approach and predictions

To assess whether responses tap into standing states, Study 3 determined the test-retest reliability of the survey from Study 2, which had already administered the survey twice to a subsample of 40 participants. **H***'s postulate of *standing* states predicts

high correlations between the ratings the same participants provide at the initial test (t1) and the retest (t2).

To assess whether the endorsed informational contents are *deployed* in judgment and reasoning tasks, we devised two philosophically relevant tasks. The *'Phenomenal Interpretation Task'* is a judgment task that asks participants to judge whether a case of hallucination qualifies as a case of 'seeing'. The dominant visual sense of 'see' carries the implication that what is seen must be around to be seen (Fischer & Engelhardt, 2020). Philosophers have discussed a subordinate 'phenomenal' sense of the verb, which is used to describe only the subject's subjective experience and does not carry this implication, so that hallucinations can qualify as cases of 'seeing' (e.g., Ayer, 1956/1990, p. 90; Jackson (1977), pp. 33–49 (Macpherson, 2013),; p. 5). Lay participants readily use this sense in judging that hallucinators 'see' the objects they hallucinate (Sant'anna & Dranseika, 2024). We reasoned that this tendency will be reinforced by IR beliefs and attenuated by DR beliefs: We assume that laypeople employ a similarity-based categorization strategy for 'seeing', which classifies instances by similarity to an associated situation schema (Rumelhart, 1978). We further assume that laypeople will regard hallucinations as in some respects similar to cases of seeing (e.g., in both cases, the subject believes they see something) and as different in others (e.g., when you hallucinate a dagger, there is no dagger around). We finally assume that laypeople will generally take cases of hallucination to involve seeing mental images. According to IR, this constitutes a similarity to cases of seeing. According to DR, this constitutes a difference. We infer that IR will reinforce the tendency to classify hallucinations as cases of 'seeing', and DR will attenuate this tendency. We thus take $\mathbf{H}^*$ to predict positive correlations between agreement ratings for IR items and agreement with the classification of hallucinations as 'seeing', and negative correlations between this and agreement ratings for DR items.

The *'Scepticism Evaluation Task'* is a reasoning task that presents participants with a scenario much used in epistemology classes (review: McKinsey, 2024): Suppose a person's brain is removed from their body, placed in a vat, and wired up to artificially receive the same stimulation it would receive in the person's everyday life. Prominent sceptical arguments infer that the subject would then enjoy the exact same experience as in their everyday life – i.e., everything will look and feel the same to them. The arguments conclude that we cannot exclude the possibility that *we* have been 'envatted'. The metaphysical and epistemological tenets of IR support the key inferences involved in this reasoning: According to *Perceptual Causal Indirect* (PCI), when a viewer sees a physical object, it causes in them a mental image of that object. Hence, the artificial stimulation of the brain-in-the-vat can cause the same mental image. Since this image determines what the world looks like to the subject, things will look indistinguishable to you when 'envatted' and when carrying on with everyday life: So you can only know what things look like to you, as per *Epistemic Indirect* (EI), and cannot exclude the possibility that things look like that to you as a result of artificial stimulation. IR thus promotes sceptical doubt and saps confidence that one can exclude the possibility of 'envatment'. By contrast, DR's metaphysical counter-claim *Perceptual Causal Direct* (PCD) prevents the sceptical reasoning from getting off the ground, while *Epistemic Direct* (ED) will reduce the plausibility of its final move. DR thus promotes anti-sceptical confidence.

Therefore, $\mathbf{H}^*$ predicts negative correlations between ratings of confidence that once can exclude the possibility of 'envatment' and agreement ratings for IR items including PCI and EI, and positive correlations between such confidence ratings and agreement ratings for DR items including PCD and ED.

2.4.2 Study 3 – Methods

Participants: 61 participants were recruited in the same way as in previous studies and completed the same survey twice, four weeks apart. One participant had to be excluded due to incomplete data, another due to failing an attention check. The remaining 59 participants were between 19 and 61 years old (mean 29.1 years, SD = 10.16). 27 were male, 30 female, 1 non-binary, and 1 'preferred not to say'.
Materials and procedure: Participants were given the same materials and tasks as in Study 2. When taking the survey for the second time (t2), however, participants received the two new tasks explained above, presented after the belief inventory. The Phenomenal Interpretation Task presented a case of hallucination:

> Suppose John is lying awake on a bed in an otherwise empty room. John has a vivid hallucination of a tiger attacking him.

Participants were then asked to rate their agreement with the statements that there is a tiger in John's room and that John sees a tiger, and to indicate their preference for one of two competing descriptions of the scenario as more accurate: 'John does not see a tiger, because this is merely a hallucination, and there is no tiger in his room' and 'John sees a tiger, but this is merely a hallucination, and there is no tiger in his room'. The Scepticism Evaluation Task presented a brain-in-a-vat scenario:

> One day, a scientist removes Alex's brain from his body, suspends it in a vat of life-sustaining liquid, and connects its neurons by wires to a supercomputer that provides the brain with electrical impulses. These electrical impulses are exactly the same that Alex's brain normally receives, as Alex goes about his daily life. Finally, the scientist erases from Alex's brain any memory of the intervention that moved his brain from his skull into the vat.

Participants then rated their agreement with the statements that Alex continues to see the same things he sees during his daily life, Alex believes he is going about his daily life, and Alex has two hands. Finally, they were asked

> Is it, for all you know, in principle possible that something like this story could have happened to you? How confident are you that you can exclude this possibility?

A 7-point scale was used for responses, and response options were presented to all participants in the same order (unlike in Study 2). Participants were asked to explain their responses to the last two tasks.

Table 4 *Studies 2–3* – test-retest correlations for verbal item ratings ($N=93$), per subscale. All p's <0.001

Perceptual Direct	Perceptual Indirect	Perceptual Causal Direct	Perceptual Causal Indirect	Phenomenal Direct	Phenomenal Indirect	Epistemic Direct	Epistemic Indirect
0.64	0.70	0.60	0.66	0.65	0.75	0.70	0.65

Table 5 Test-retest correlations for animation agreement ratings by correct interpreters (CIs)

Direct animation			Cartesian animation		
CIs: N (% of sample)	r	p	CIs: N (% of sample)	r	p
10 (10.6)	0.86	<0.001	33 (35.3)	0.60	<0.001

Analysis: To assess test-retest reliability and the stability of the beliefs of interest, we examined correlations between verbal and image ratings provided upon first and second trial, respectively, both for the present sample of 59 and for a sample of 99 obtained by including data from the 40 participants in Study 2, who had taken the same questionnaire twice. Prior to the analysis we assessed multivariate outliers, across the 8 subscales and 2 time points. We classed values > 2.5 SDs from the mean as outliers. 3 participants exceeded this criterion (i.e., greatly exceeded the mean discrepancy between t1 and t2 ratings of the overall sample) and were excluded, as were 3 participants whose responses at t2 were strategic or incomplete.

To assess the inferential deployment of these beliefs, we examined correlations between ratings of the items in our 8 subscales and responses to the two new tasks.

2.4.3 Study 3 – Results

Test-retest correlations for all eight item categories were significant (all p's <0.001). For psychometric questionnaires, Cicchetti (1994) defined correlations of 0.4–0.59, 0.6–0.74, and ≥0.75, respectively, as indicating fair, good, and excellent test-retest reliability, respectively. By these lights, all our eight conditions showed at least good reliability (Table 4).

We also calculated test-retest correlations for animation ratings, after excluding 4 participants as outliers and one due to incomplete data ($N=94$; all p's <0.001): Cogs $r=0.47$; Screen $r=0.35$, Cartesian $r=0.55$, Direct $r=0.76$. Since the proportions of participant who interpreted these animations correctly were only between 20 and 60%, we also calculated test-retest correlations between ratings for the key Direct and Cartesian animations, just for those participants who interpreted the image correctly both times they took the survey (Table 5). Appendix D provides information about the test-retest reliability of compatibility ratings.

Inferential deployment: After excluding 2 participants due to incomplete data, we examined whether agreement with DR and IR statements and animations (at t2) correlates in different directions with responses to the Phenomenal Interpretation Task and the Scepticism Evaluation Task (Sect. 2.3.1). In the former task, we found no significant correlations for any subscale with agreement ratings for 'John sees a tiger' (ED $p=0.12$, all other p's >0.28, 2-tailed) or with preferences for the competing descriptions denying vs affirming 'seeing' (all p's >0.52). By contrast, in the

Table 6 Study 3 – Scepticism evaluation task ($N=59$): correlations between ratings of confidence that one can exclude the sceptical scenario and agreement ratings in eight verbal and one animation condition

	PD	PI	PCD	PCI	PHD	PHI	ED	EI	Cartesian
r	0.27	−0.27	0.32	−0.31	0.29	−0.21	0.27	−0.44	−0.22
p (2-tailed)	0.04	0.04	0.01	0.02	0.02	0.11	0.04	<0.001	0.1

Scepticism Evaluation Task, we found significant medium-sized correlations in the predicted directions between participants' confidence to be able to exclude they are in the sceptical (BIV) scenario and their agreement with Perceptual Causal Direct, Perceptual Causal Indirect, and Epistemic Indirect (Table 6). We found further significant but small correlations in the predicted direction with Perceptual Direct, Perceptual Indirect, Phenomenal Direct, and Epistemic Direct.

2.4.4 Study 3 – Discussion

Test-retest findings support both questionnaire development and validation. Factor analyses had revealed that one set of metaphysical conditions is surplus to requirements (Sect. 2.3.3). The higher test-retest reliability of the Perceptual than the Perceptual Causal conditions suggests retaining the former and dropping the latter. For validation, the key question is whether good test-retest reliability is good enough to support the hypothesis **H*** of an underlying standing state. Given the narrowness of the concept, one might well have expected excellent test-retest reliability. However, Studies 1–2 revealed that a sizable proportion of participants felt torn between conflicting responses and that well over half of these participants had insight into the conflicting nature of these responses (see Appendix D). Taking the initial test (including the compatibility assessment) will lead such participants to grow aware of their potential belief conflicts. Standard views about cognitive dissonance would predict that many of these participants – and certainly all those with high levels of preference for consistency (Cialdini et al., 1995) – will 'correct' their responses at the retest, attenuating test-retest correlations. These would then reflect not measurement error, but changes in the 'true value', which were only induced by the study task. We conclude that good test-retest reliability is good enough to provide evidence that the observed ratings are due to a standing state.

We consider next potential evidence that this state influences judgment and reasoning tasks. Findings from the judgment task (Phenomenal Interpretation) disconfirmed our prediction. To confirm this outcome, we considered further results. Two thirds of participants agreed that the hallucinator, John, sees a tiger (while 12% remained agnostic and 22% disagreed). This is in line with the high proportions of similar responses observed in previous studies (Sant'anna & Dranseika, 2024), but too low to make ceiling effects a plausible explanation of the observed lack of correlations between task-specific agreement ratings and ratings for DR and IR items. This lack of correlations was underscored by a manual cluster analysis that classified participants as 'Direct Realists' and 'Indirect Realists', respectively, based on their mean agreement ratings for PD and PI (numerically above or below mid-point 4, respectively): Virtually the same proportions of 'pure' Direct and Indirect Realists agreed that John sees a tiger (71.4% vs 69.2%), disagreed (20% vs 23.1%), or

remained agnostic (8.6% vs 7.7%). Those metaphysical beliefs about vision seem to have made no difference to agreement ratings for 'John sees a tiger'. This conclusion was confirmed by analysis of participants' free-text explanations of their responses (see Appendix E).

By contrast, findings from the reasoning task (Scepticism Evaluation) confirmed predictions. We identified medium-sized correlations in the predicted directions between scepticism evaluations and agreement ratings for PCI, EI, and PCD, i.e., three of the four beliefs we identified as supporting and preventing the sceptical reasoning, respectively. For the fourth (ED), we observed a small-to-medium correlation. The further correlations observed largely follow from the correlations between item ratings previously observed in Studies 1–2 (see Appendix C, Table C2).

To sum up, findings from the Phenomenal Interpretation Task suggest that differences between DR and IR beliefs make no difference to classification judgments of hallucination cases as instances of 'seeing'. The Scepticism Evaluation Task provided evidence that these beliefs influence reasoning about a hypothetical situation. We infer that different knowledge structures are deployed for these tasks. Classification judgments deploy conceptual information about the category at issue. A *concept* is a body of information stored in long-term memory that is retrieved by default, in the exercise of higher cognitive competencies including language comprehension, perceptual and cognitive categorisation, and inductive learning (review: Machery, 2009). The present classification judgments will deploy the concept associated with the verb 'to see', a situation schema that encodes information about the typical and diagnostic features of agents (viewers) and patients (objects of sight), and typical causal and other relations between them (Hampton, 2006; Rumelhart, 1978), let's call it the '*seeing*-schema'. Findings from the Phenomenal Interpretation Task then suggest that Direct Realists and Indirect Realists share the same concept of *seeing*. If so, it is *not* the case that the *seeing*-schema of Direct Realists includes DR beliefs, while the *seeing*-schema of Indirect Realists includes IR beliefs. Below (Sect. 3), we will suggest that DR statements are reflective of the situation schema associated with 'see'. We infer that IR beliefs are part of different knowledge structures, which are not deployed in classification judgments but in reasoning about hypothetical situations.

Together, Studies 1–3 support the development of the new Direct/Indirect Realist Belief Inventory (DIRBI) and provide first evidence for the hypothesis **H*** that the agreement ratings elicited by this new measure are indicative of genuine *beliefs*: Present findings suggest that the DIRBI has the requisite structural validity (Studies 1–2) and taps into information-bearing states that are standing (Studies 2–3) and inferentially deployed (Study 3). We can therefore employ findings from these DIRBI studies to assess our three hypotheses about conflicting beliefs.

2.5 Further analyses: assessing H_1-H_3

To assess our hypotheses that laypeople are ($\mathbf{H_1}$) collectively and ($\mathbf{H_2}$) often individually torn between DR and IR conceptions of vision, which ($\mathbf{H_3}$) are stored in distinct belief structures ('fragments'), we now return to Studies 1–2 and report further analyses of their data, which assess our predictions from these three hypotheses.

2.5.1 Assessing H_1-H_2 (studies 1–2)

$\mathbf{H_1}$ predicts that a higher proportion of lay participants than can be plausibly explained by performance error will endorse DR and IR, respectively. $\mathbf{H_2}$ predicts that a non-negligible proportion of laypeople will endorse both DR and IR conceptions simultaneously. To assess $\mathbf{H_1}$-$\mathbf{H_2}$, cluster analyses sought to determine the proportions of participants in Studies 1–2 who adhere to DR, to IR, to both ('Torn Souls'), and to neither ('Refuseniks'). These analyses were informed by previous factor analyses. Exploratory factor analyses revealed that all DR conditions load positively on one factor, and all IR conditions load on another. This finding suggests that it makes sense to assess adherence to DR and IR by considering an individual's mean agreement ratings across all DR conditions and across all IR conditions, respectively. To do so, we used the saved factor scores from the exploratory factor analyses. The results of the confirmatory factor analysis provide a positive reason to do so: They expose some dissociation between metaphysical and epistemological beliefs of each kind. This finding suggests that we cannot single out any one condition as being indicative of the DR or IR nature of the individual's conception of vision, and need to consider means across all conditions.

In both Study 1 and Study 2, we first conducted a data-driven cluster analysis, viz., a k-means analysis based on the two relevant factors 1 (*'Direct Factor'*) and 2 (*'Indirect Factor'*). To do so, we used in each study the saved factor scores generated by the prior exploratory factor analysis (reported in Sects. 2.2.2 and 2.3.2, respectively) and instructed the SPSS analysis software to identify four groups, hoping that these would correspond to the four groupings sought. In both studies, the centroids of the four groups indeed fell into the four quadrants of the relevant coordinate space (see Appendix F) and permitted similarity-based classification of participants as Direct Realists, Indirect Realists, Torn Souls, and Refuseniks, namely, as adherents to the position to whose prototypical ratings (as indicated by centroids) their scores are most similar.

We complemented the data-driven analysis with two manual cluster analyses that are more straightforward to interpret. The *Threshold* analysis classified a participant as 'Direct Realist' if and only if their mean rating for items positively loading on the Direct Factor was numerically above mid-point '4' and their mean rating for items positively loading on the Indirect Factor was ≤ 4. The converse holds for 'Indirect Realists'. For both item classes, 'Torn Souls' had means numerically above 4, and 'Refuseniks' means ≤ 4. The *Discrepancy* criterion sought to capture marked preferences that might translate into differences in how DR and IR conceptions are cognitively deployed. This criterion insisted on a difference of at least 1 point between mean agreement ratings for DR and IR items (on the 7-point scale employed in Studies 1–2). If their ratings for DR items were at least 1 point higher than for IR items, participants were classified as 'Direct Realists'; in the reverse case as 'Indirect Realists'. Where mean ratings differed by less than one point between these two kinds of items, participants were classified as 'Torn Souls' if both means were above mid-point, and as 'Refuseniks' if both means were below mid-point. Three participants with means on both sides of, or just at the mid-point, were classified as 'Neutral'. Findings from these cluster analyses are visualised by Fig. 4 and reported in Table 7.

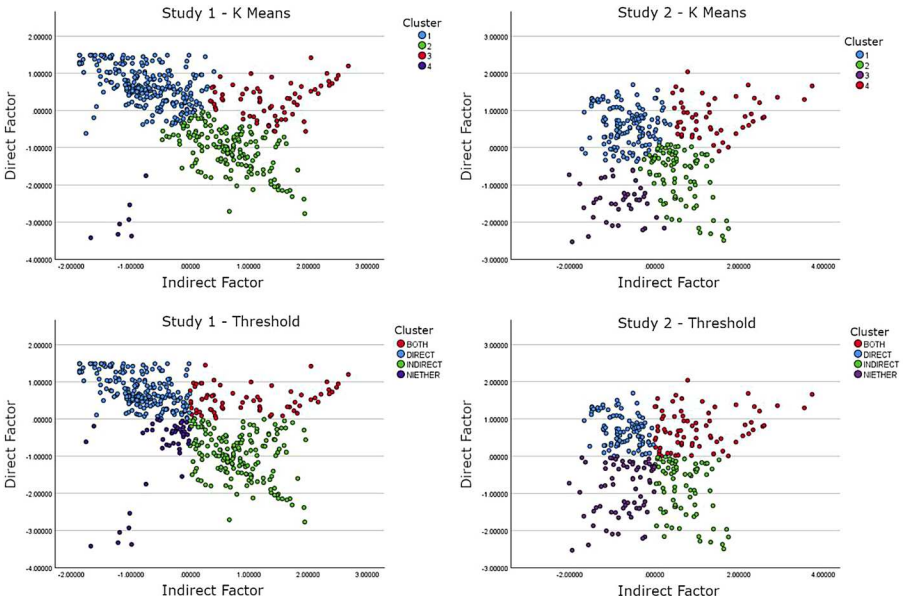

Fig. 4 Studies 1 and 2 – scatterplots showing classification of participants with k-means analysis (upper row) and by Threshold criterion (lower row) in Study 1 and Study 2. X-axes show factor scores on the 'Indirect factor'. Y-axes show factor scores on the 'Direct factor'. All graphs use the same colour code: In both studies, cluster '1' identified by k-means analysis is interpreted as grouping together the Direct Realists (blue), 'cluster 2' the Indirect Realists (green), '3' the torn souls ('both'; red), and '4' the Refuseniks ('neither'; purple)

Table 7 Studies 1 and 2 – number of individuals (% of sample) classified as Direct Realist, Indirect Realist, torn soul ('both') and refusenik ('neither'), by different cluster analyses. Bottom lines indicate the ratios of Direct Realists (DR) to Indirect Realists (IR), and of participants classified as Direct Realist or both (DR*) to participants classified as Indirect Realist or both (IR*), by the different analyses: k-means, threshold and discrepancy criterion

	Study 1:			*Study 2*:		
	K-means	Threshold	Discrepancy	K-means	Threshold	Discrepancy
Direct	261 (52.2)	215 (43.0)	174 (34.8)	138 (43.3)	100 (31.3)	90 (28.2)
Indirect	164 (32.8)	172 (34.4)	172 (34.4)	90 (28.2)	75 (23.5)	78 (24.5)
Both	68 (13.6)	67 (13.4)	81 (16.2)	51 (16.0)	74 (23.3)	82 (25.7)
Neither	7 (1.4)	46 (9.2)	62 (12.4)	40 (12.5)	70 (21.9)	69 (21.6)
DR:IR	1.59:1	1.25:1	1.01:1	1.53:1	1.33:1	1.15:1
DR*:IR*	1.42:1	1.18:1	1.01:1	1.34:1	1.17:1	1.08:1

Note: Discrepancy analyses created 3 'neutral' participants in Study 1

We next analysed responses to the **animations** for Study 1 and Study 2. Of particular interest were responses to the Direct and Cartesian animations, which represent the paradigmatic forms of Direct and Indirect Realism, respectively. For each image, we considered what proportion of participants interpreted the images as intended, and what proportion of these 'correct interpreters' agreed that the image 'accurately represents what happens when we see an apple' (Table 8).

Table 8 Studies 1 and 2 – number of participants who correctly interpret Direct and Cartesian animations ('correct interpreters') and also agree with animations ('competent agreement'). In parentheses, proportions of whole sample and of correct interpreters, respectively

	Direct animation		Cartesian animation	
	Study 1 (N = 500)	Study 2 (N = 319)	Study 1 (N = 500)	Study 2 (N = 319)
Correct interpreters	134 (26.8%)	61 (19.1%)	297 (59.4%)	170 (53.3%)
Competent agreement	78 (58.2%)	35 (57.3%)	140 (47.1%)	76 (44.7%)

Table 9 Studies 1 and 2 – classification of participants who competently agree with the Direct and the Cartesian animation, respectively: classification by verbal agreement ratings, with Threshold criterion (T) and discrepancy criterion (D)

N = competently agree with item	Direct animation				Cartesian animation			
	Study 1 (N = 78)		Study 2 (N = 35)		Study 1 (N = 140)		Study 2 (N = 76)	
Direct Realists	T 46	D 37	T 19	D 16	T 40	D 29	T 15	D 10
Indirect Realists	T 16	D 13	T 3	D 3	T 71	D 70	T 25	D 28
Torn Souls (Both)	T 10	D 16	T 8	D 11	T 21	D 24	T 13	D 20
Refuseniks (Neither)	T 6	D 12	T 5	D 5	T 8	D 17	T 23	D 18

Responses to verbal items and to animations may diverge (as predicted by the fragmentation accounts of belief storage to which we will return below). For each of the two key animations, we therefore considered how those participants who agreed that the animation 'accurately represents what happens when we see an apple' had responded to verbal items. More specifically, we considered how these participants had been classified by the manual cluster analyses conducted on verbal agreement ratings (Table 9). This will allow us to better gauge the proportion of participants who feel torn between the DR and IR conceptions.

We finally analysed responses to the **compatibility** questions in Studies 1–2 and found that participants, including a strong majority of 'Torn Souls', regarded DR items (PD, PHD, and ED) as incompatible with their IR counterparts (PI, PHI, EI) (Appendix D).

Findings from cluster and further analyses allow us to assess hypotheses $\mathbf{H_1}$-$\mathbf{H_2}$. $\mathbf{H_1}$ is most directly assessed by the cluster analyses of agreement ratings for verbal items. $\mathbf{H_1}$ states that laypeople are collectively torn between DR and IR and predicts that a higher proportion of lay participants than can be plausibly explained by performance error will endorse DR and IR, respectively. To assess $\mathbf{H_1}$, we consider the ratio of Direct Realists to Indirect Realists identified by those cluster analyses (DR:IR in Table 7). Across both studies, the manual analyses that are the most straightforward to interpret yield ratios between 1.33:1 and 1.01:1. The slight preponderance of Direct Realists is further attenuated the moment we also take into account participants who endorse both views and consider the ratio of participants who endorse only *or also* DR to those who endorse only *or also* IR (DR*:IR* in Table 7). Both conceptions are popular and seem to be roughly equally popular, so that agreement with neither conception can be plausibly attributed to performance error. $\mathbf{H_2}$ suggests that many laypeople are individually torn between DR and IR. To assess this suggestion, we consider first the proportion of participants who endorse both conceptions, when

stated verbally ('Both' in Table 7). Depending upon the classification and sample used, this proportion varies between 13 and 26%. To these participants we can add those who endorse verbal statements only of DR but also understand and agree with the IR Cartesian animation (Table 9). In Studies 1 and 2, respectively, these are 8.0% and 4.7% of the overall sample (and 28.6% and 19.7% of correct interpreters of the animation). Thus 21–31% of participants provide evidence of agreeing with both DR and IR. Compatibility findings reveal most of them do so in awareness of the conflict between the two conceptions. We infer that relevant responses cannot credibly be attributed to performance errors and reveal a non-negligible proportion of torn souls. Present findings support both $\mathbf{H_1}$ and $\mathbf{H_2}$.

2.5.2 Assessing H_3 (studies 1–2)

The explanatory hypothesis $\mathbf{H_3}$ seeks to explain the inter- and intrapersonal disagreements identified by $\mathbf{H_1}$-$\mathbf{H_2}$ by suggesting that laypeople can draw on two internally coherent conceptions of vision that are stored in different belief fragments. On the most straightforward interpretation, each of the relevant memory stores contains a comprehensive and internally coherent conception of vision that comprises both the metaphysical and the epistemological beliefs of DR and IR, respectively. On this interpretation, $\mathbf{H_3}$ predicts that ratings for DR and IR items should positively load on different factors. By contrast, the orthodox view that laypeople widely hold one conception of vision, which is consistent with DR, would predict that ratings for all items are due to the same construct, so that all subscales should load on the same factor, DR subscales positively, and IR subscales negatively.

In both Studies 1 and 2, and in follow-up analysis of their merged data, exploratory factor analysis identified two factors: one on which all DR items load positively, and one on which all IR items load positively. This speaks against the orthodox view and for our hypotheses. The results of our confirmatory factor analyses, however, complicate the picture: The only model with good fit was achieved in our follow-up analysis of the merged data from Studies 1 and 2 (Sect. 2.3.3) and assigned Metaphysical and Epistemological Direct items to different factors, and could not accommodate Epistemological Indirect items. On an account consistent with $\mathbf{H_3}$, this could be due to the fact that agreement ratings are influenced across the board by both DR and IR beliefs, whose relative influence varies between subscales.

To further investigate these dissociations, we manually classified all our 819 participants in Studies 1–2 by the consistency of their ratings: 75% either agreed or disagreed consistently with all of Perceptual, Phenomenal, and Epistemic Direct, as evidenced by mean ratings for each of these conditions numerically above or below mid-value 4, respectively. 16.1% agreed with two of these three beliefs, while 8.9% agreed with only one of them. By contrast, only 59.1% agreed or disagreed consistently with all of Perceptual, Phenomenal, and Epistemic Indirect. 15.8% agreed with two of these three beliefs, while 25.2% agreed with only one of these beliefs. That is, while three quarters of participants had a consistent attitude to DR, about 40% of participants had conflicting attitudes to IR, and accepted some of its constituent beliefs, while rejecting others. This suggests that at any rate the different component

beliefs of IR are not generally stored in one belief fragment, whose contents would be typically activated or suppressed together, but are 'massively fragmented'.

In addition, present analyses provide support for two key ideas of the fragmentation accounts of belief storage that inform $\mathbf{H_3}$, namely, (i) the idea that one and the same individual may store conflicting beliefs, and (ii) the idea that these beliefs get activated independently, by different stimuli or in different contexts. Compatibility ratings (Sect. 2.5.1) reveal that well over half of the participants who explicitly endorse verbal statements of DR and IR deem these statements incompatible. This supports (i). To assess (ii), we considered the consistency of participants' agreement ratings across different stimulus formats. Fischer et al. (2023) had observed that pictorial stimuli are more liable than verbal stimuli to elicit agreement with IR. We therefore considered to what extent the Cartesian animation attracted competent agreement from participants qualifying as Direct Realists, by verbal ratings: depending upon the classification criterion used, 16.7–18.6% of these participants competently agreed with the Cartesian animation in Study 1, and 11–15% did so in Study 2. This is consistent with the suggestion that different stimulus formats preferentially activate different conceptions.

3 General discussion

Three studies sought to validate a new measure of beliefs about vision, the *Direct/Indirect Realist Belief Inventory (DIRBI)*, that was specifically designed to assess conflicting Direct Realist and Indirect Realist beliefs (and is described in detail in Appendices A) and B and on OSF). Findings suggest that the agreement ratings elicited with the DIRBI are indicative of the relevant beliefs. We further employed the new measure (1) to assess inter- and intrapersonal conflicts between competing Direct Realist (DR) and Indirect Realist (IR) conceptions of vision and (2) to examine whether these are stored in different belief fragments. In philosophy, a widespread consensus takes for granted that there is such a thing as 'the' folk conception of vision, and that this is consistent with DR and inconsistent with IR. Our new measure proved its value by documenting, against this orthodoxy, conflicts between DR and IR folk conceptions, between subjects (as per $\mathbf{H_1}$) and within subjects (as per $\mathbf{H_2}$), and providing evidence of fragmentation in folk beliefs about vision (as per $\mathbf{H_3}$). Findings suggest 'massive fragmentation', where fragments do not store comprehensive DR and IR conceptions of vision, each comprising both metaphysical and epistemological beliefs, but smaller belief sets.

We now discuss our main findings concerning the validation of the DIRBI and the knowledge structures this new inventory taps into. This discussion will build up to an explanation of why competent thinkers may maintain mutually incompatible DR and IR beliefs about vision. As noted in the Introduction, plausible mid-strength principles of charity postulate that we can attribute conflicting beliefs to competent thinkers only if we can empirically explain why they hold such beliefs (Thagard & Nisbett, 1983). The proposed explanation will thereby complete our case that laypeople are not only collectively but often also individually, torn between conflicting beliefs about vision.

To show that responses to the DIRBI inventory are indicative of beliefs, understood as accessible, standing, information-bearing states that are endorsed as true and are inferentially deployed, e.g., in reasoning tasks, we need to assess this hypothesis ($\mathbf{H}^*$ above) against the alternatives that those responses are due to (1) *acquiescence bias* (Jackman, 1973) or (2) *social desirability bias* (Phillips & Clancy, 1972), or (3) are purely ad hoc (i.e., reflect potentially short-lived opinions formed while taking the survey, rather than standing beliefs).

The stable factor structure across Studies 1–2 suggests that when different people respond to the same stimuli, on different occasions, the same constructs influence responses. Good to excellent test-retest reliability in Study 3 suggests the same constructs influence responses of the same individual at different occasions. These findings are consistent with $\mathbf{H}^*$ and speak against the alternative hypothesis that responses were purely ad hoc. To guard against acquiescence bias, we provided response options in different order to different halves of our samples in Studies 1–2. Acquiescence bias would predict order effects, which we did not observe. In the context of a scientific survey on the pleasingly apolitical topic of vision, social desirability bias would most likely materialise as a bias towards agreeing with responses deemed scientific. However, a scientific accuracy rating study with participants recruited with similar restrictions as our Studies 1–3, found against the assumption that DR is deemed naïve and IR scientific (Fischer et al., 2026): When speaking abstractly of seeing 'physical objects', neither DR nor IR statements attracted mean scientific accuracy ratings significantly different from neutral midpoint (or each other). In the concrete formulation used in our Studies 1–3, IR attracted mean scientific accuracy ratings just below midpoint and DR mean ratings just above midpoint, but the latter comparison was driven by a single item (the epistemological statement ED). In addition, participants reported on average that it was merely 'slightly important' to them that their beliefs are scientifically accurate, further speaking against the relevance of social desirability bias. We can thus set aside the three most salient confounds.

Principal factor analysis of subscales found factor loadings suggesting the four items on each subscale tap into the same construct, further supported by high internal consistency (Cronbach's $\alpha = 0.82\text{-}0.89$ across subscales). This is consistent with the suggestion that each subscale probes the same 'core belief'. Stable factor structure (Studies 1–2) and good test-retest reliability ($r = 0.60\text{-}0.75$ across subscales) (Study 3) further support the hypothesis $\mathbf{H}^*$ that responses are due to accessible and standing information-bearing states. A reasoning task provided first evidence that their information is inferentially deployed (Study 3). We infer the DIRBI taps into beliefs of some sort.

Present findings suggest interesting hypotheses concerning the structure and nature of the relevant knowledge structures. Exploratory factor analysis in Studies 1–2 (and on merged samples) revealed two factors: agreement ratings for DR subscales all loaded positively on one factor, on which metaphysical IR subscales loaded significantly, but negatively. This suggests we are dealing with at least two different belief structures: One set of DR beliefs influenced responses to both DR items (positively) and at any rate metaphysical IR items (negatively), while a set of IR beliefs additionally influenced responses to IR items (positively). Confirmatory factor analysis revealed further structure within these belief sets, dissociating metaphysical and

epistemological beliefs. Manual classification provided further evidence that metaphysical and epistemological beliefs do not form coherent belief sets, in particular on the IR side: Whereas 75% of participants consistently accepted or rejected all DR beliefs and a mere 9% agreed with just one of the DR beliefs probed, only 60% of participants had a similarly consistent attitude towards IR beliefs, with 25% accepting only one of the IR beliefs probed.

Taken together, these results suggest that the probed DR beliefs are due to two intertwined belief structures that may share a common source, whereas the metaphysical, phenomenal, and epistemic core beliefs of IR are part of distinct belief structures, or fail to be integrated into larger structures. These findings are mirrored by the structures of philosophical theories of vision: Whereas DR theories tend to combine all probed metaphysical and epistemological claims, only some IR theorists have endorsed PI, PHI and EI (e.g., Russell, 1912/2000), whereas others have been more selective (e.g., Jackson, 1977; see; Lyons, 2023 for discussion). In brief, we are not looking at comprehensive DR and IR folk theories of vision, which combine metaphysical and epistemological beliefs, but at smaller knowledge structures.

So what are these knowledge structures? DR and IR are often regarded as naïve and scientific theories of vision, respectively (Sect. 1.2). This would predict more agreement with IR among participants with higher education (A-level or above) in natural sciences, psychology, or philosophy. But we observed only negligible (Study 1) or no effects (Study 2) of such higher education. Moreover, a parallel scientific accuracy rating study revealed that, unlike Newtonian claims about colour, neither DR nor IR is recognised by laypeople as a scientific theory, or even treated as a unified account to be comprehensively assessed for scientificality (Fischer et al., 2026).

We simultaneously have more fundamental doubts about regarding either account as a *theory*, in the sense of a *causal-explanatory framework*. The positive claims of DR (e.g., 'When you look at an orange, you see just the orange …') derive their content largely from the contrast with opposing IR beliefs ('… and not a mental image of the orange'), in marked contrast to the causal-explanatory claims of exemplary naive theories (Gelman & Noles, 2011; Vosniadou, 1994). Philosophers have used IR to explain 'facts' about illusions and hallucinations (e.g., Jackson, 1977; Robinson, 1994), so IR beliefs could form part of naive theories of these specific phenomena, if unsuited to contribute to a more general naïve theory of vision. However, the rationales for seeking the pertinent explanations of illusions rely on substantive philosophical assumptions that include the 'Phenomenal Principle' ('Whenever something appears F, the viewer is aware of something that *is* F') that, in a lay sample, Fischer et al. (2023) found not to be associated with IR beliefs. Similarly, open-text responses to the Phenomenal Interpretation Task in Study 3 revealed that many laypeople invoke mental images to conceptualise hallucinations, but that the probed DR and IR beliefs about (regular cases of) vision do not influence the propensity to do so (Appendix E). This suggests that the probed IR beliefs about vision do not figure in a naïve theory of hallucination. It thus remains unclear what, if anything, such folk beliefs would be used to causally explain, in lay cognition.

We suggest instead that IR beliefs about vision are an artefact of an implicit model of selective attention. According to 'attention schema theory' (Graziano, 2022; Graziano & Kastner, 2011), processes of social cognition compute implicit perceptual

models of attention that help people track others' focus of attention, support endogenous control of selective attention, and are partly accessible to higher cognition, giving rise to intuitions and beliefs. Implicit models prioritize usefulness and economy over accuracy, and provide useful caricatures, rather than accurate representations of the process they depict. Thus, for the purpose of tracking others' visual attention, processes of social cognition compute a perceptual model that assigns a set of attributes to spatial locations: The model ties an attention signal to a place (a subject of attention), directs it towards a spatially located object (of attention), and represents direction of attention by a subthreshold motion signal that indicates a flow from the viewer's eyes to the object (Guterstam & Graziano, 2020). Partial access of higher cognition to this implicit model could explain the extramissionist belief that vision involves a force ray travelling from eyes to object (Guterstam et al., 2019).

Where the model constructed is of one's own, rather than another's attention, it ties the attention signal to a place in one's own head. Since attention can be directed not only towards external objects but also towards one's own thoughts, beliefs, feelings, sensations, etc., the model represents our attention as directed towards objects spatially located in our head (thoughts, beliefs) or body (corporeal sensations). Graziano (2022) seeks to explain common intuitions and beliefs about consciousness by reference to this perceptual model. We suggest partial access of higher cognition to this implicit model might explain belief in the Cartesian Theatre conception and metaphysical IR beliefs about vision embedded in that conception. Since the implicit model of attention does not represent any inferences, this explanation might also account for the dissociation of metaphysical and epistemological IR beliefs.

We further suggest another implicit knowledge structure is the ultimate source of DR beliefs. Verbal and visual stimuli automatically activate concepts. Prototype concepts organise experiential knowledge built up through observation of the physical and discourse environment (Rosch & Mervis, 1975; review:; Hampton, 2006). Verbs (like 'to see') automatically activate *situation schemas* that encode internally structured experiential event knowledge about the typical and diagnostic features of agents (viewers) and patients (objects of sight), and typical relations between them (McRae et al., 1997; Rumelhart, 1978). Dependency networks in complex schemas can encode causal, functional, and nomological information (Sloman et al., 1998). Such conceptual knowledge about *seeing*-events will influence responses in our present studies. Vignette-based studies provide evidence that the *seeing*-schema includes the knowledge that the viewer's line of sight must not be blocked by another object, but no knowledge of causal mechanisms in line with IR (Roberts et al., 2018). While implicit, conceptual knowledge organized by prototypes and situation schemas is accessible to processes of higher cognition. We suggest that DR beliefs are gradually formed as explicitly stored beliefs when this experiential event knowledge is deployed to assess IR beliefs, as and when these are activated on different occasions. Once formed, the DR beliefs are stored in long-term memory. This account could explain the contrastive character of DR beliefs ('When you look at an orange, you see just the orange…' – 'just', instead of what?), which seem to depend for their content on the contrast with IR beliefs ('… and not a mental image of an orange').

Since the implicit knowledge structures we hypothesize to be at the root of the relevant beliefs are well-nigh universal, the variability we observed in responses must

be due to individual differences downstream: they would need to arise from individual differences in conflict monitoring and detection, or in conflict management. While the impact of individual difference variables has been extensively studied for conflicts between belief- and rule-based responses in judgment and decision tasks (reviews: Erceg et al., 2022; Šrol & De Neys, 2020), the influence of individual difference variables affecting the detection of, and attitude towards conflicts between own beliefs remains underexamined (Musolino, Sommer, & Hemmer, 2022). Further discussion is reserved for another occasion.

To sum up, we interpret the observed conflict between DR and IR beliefs as a conflict between competing pre-scientific conceptions ultimately grounded in experiential event knowledge and an implicit model of endogenous attention, respectively. Together with findings that both DR and IR beliefs are treated as naïve in lay cognition (Fischer et al., 2026), present findings support the conclusion that mutually incompatible conceptions of vision compete *within* common sense, so that there is no such thing as 'the' (unified, coherent) common-sense conception of vision.

4 Conclusion

To study folk beliefs of philosophical interest, and especially conflicting folk beliefs that give rise to aporetic problems, we need to address the philosophically neglected task of developing and validating belief inventories. A new belief inventory (DIRBI) allows to probe Direct Realist (DR) and Indirect Realist (IR) beliefs about vision. (This inventory is made available in Appendices A–B and on OSF.) A validation study provided first evidence that the new measure manages to deploy agreement ratings to tap into beliefs, understood as accessible, standing, information-bearing states that are endorsed as true and are inferentially deployed, e.g., in reasoning tasks. The new measure provides evidence that laypeople are, collectively and often individually, torn between conflicting beliefs about vision, consistent with DR and IR, respectively. These beliefs may be co-represented in the same individual, across different belief fragments that are activated by different stimuli. We argued these belief structures are ultimately rooted in two different kinds of partially accessible implicit knowledge structures, namely, experiential event knowledge organised by the situation schema associated with the verb 'to see' (DR), and an implicit model of attention used to control endogenous selective attention (IR). Present findings have major methodological implications for debates about the nature of perception and the aporetic 'problem of perception', which accord epistemic default status to the common-sense conception of vision: We found there is no such thing as 'the' common-sense conception that could enjoy this status. This raises pressing methodological questions about longstanding debates concerning other aporetic problems.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11229-026-05575-0.

Acknowledgements We thank Nina Haket for assistance with data analysis and data archiving. For helpful comments on previous drafts we are grateful to Joshua Knobe, Gavin Nobes, and audiences at the 3rd European Experimental Philosophy conference, the 2024 conference of the European Society for Philosophy

and Psychology, the inaugural meeting of the UK Experimental Philosophy Triangle, the British Academy, and the Universities of Miami and Nottingham, as well as two anonymous reviewers.

Author contributions All authors contributed equally to conceptualisation, study design, material development, and interpretation of findings. Paul Engelhardt analysed and curated the data. Eugen Fischer wrote the original draft of the paper, which was reviewed and edited by all authors. Fischer and Keith Allen acquired funding from the British Academy, Fischer and Engelhardt acquired funding from the Thyssen Foundation.

Funding This research was funded by the British Academy (SRG22\220454; Studies 1-2) and the Fritz Thyssen Foundation (10.22.2.010PH; Study 3).

Data availability Materials and data files are available through OSF: https://osf.io/3mcep/overview

Declarations

Ethics approval The use of human research participants was approved by the Arts and Humanities Research Ethics Committee of the University of York. The research conformed to the ethical standards for conducting research as outlined by the British Psychological Society.

Conflict of interest None of the authors is aware of any conflicts of interest.

References

Albarracin, M., & Pitliya, R. J. (2022). The nature of beliefs and believing. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.981925

Allen, K. (2019). Merleau-Ponty and Naïve realism. *Philosophers' Imprint*, *19*(2), 1–25.

Allen, K. (2020). The value of perception. *Philosophy and Phenomenological Research*, *100*(3), 633–656. https://doi.org/10.1111/phpr.12574

Aristotle (n.d.). R. Crisp (Ed.). *Nicomachean Ethics*. CUP.

Austin, J. L. (1962). *Sense and Sensibilia*. OUP.

Ayer, A. J. (1956/1990). *The problem of knowledge*. Penguin.

Barlev, M., Mermelstein, S., & German, T. C. (2018). Representational coexistence in the God concept: Core knowledge intuitions of God as a person are not revised by Christian theology despite lifelong experience. *Psychonomic Bulletin & Review*, *25*(6), 2330–2338. https://doi.org/10.3758/s13423-017-1421-6

Bendaña, J. &., & Mandelbaum, E. (2021). The fragmentation of belief. In C. Borgoni, D. Kindermann, & A. Onofri (Eds.), *The fragmented mind* (pp. 78–107). OUP.

Berkeley, G. (1709/1975). An essay towards a new theory of vision. In *Philosophical works* (pp. 1–70). J.M. Dent.

Bertossa, F., Besa, M., Ferrari, R., & Ferri, F. (2008). Point zero: A phenomenological inquiry into the seat of consciousness. *Perceptual and Motor Skills*, *107*(2), 323–335. https://doi.org/10.2466/pms.107.2.323-335

Boghossian, P., & Velleman, D. (1989). Colour as a secondary quality 1. *Mind*, *98*(389), 81–103. https://doi.org/10.1093/mind/XCVIII.389.81
Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard UP.
Burge, T. (2005). Disjunctivism and perceptual psychology. *Philosophical Topics*, *33*(1), 1–78. https://doi.org/10.5840/philtopics20053311
Carey, S. (2009). *The origin of concepts*. OUP.
Cath, Y. (2016). Reflective equilibrium. In H. Cappelen, T. Szabó Gendler, & J. Hawthorne (Eds.), *The Oxford handbook of philosophical methodology* (pp. 213–230). OUP.
Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276.
Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality & Social Psychology*, *69*(2), 318–328. https://doi.org/10.1037/0022-3514.69.2.318
Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284
Crane, T., & French, C. (2021). The problem of perception. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/fall2021/entries/perception-problem
Crimmins, M. (1992). Tacitness and virtual beliefs. *Mind & Language*, *7*(3), 240–263. https://doi.org/10.1111/j.1468-0017.1992.tb00207.x
Davia, C. (2017). Aristotle and the endoxic method. *Journal of the History of Philosophy*, *55*(3), 383–405. https://doi.org/10.1353/hph.2017.0047
Dennett, D. C. (1991). *Consciousness explained*. Little, Brown.
Descartes, R. (1641/1996). Cottingham, J. *Meditations on first Philosophy with selections from the objections and replies*. CUP.
Drayson, Z. (2021). Naturalism and the metaphysics of perception. In H. Logue & L. Richardson (Eds.), *Purpose and procedure in the philosophy of perception*. Oxford University Press.
Dretske, F. (1988). *Explaining behavior*. MIT Press.
Dundes, A. (1981). *The evil eye: A folklore casebook*. Garland.
Erceg, N., Galić, Z., Bubić, A., & Jelić, D. (2022). Who detects and why: How do individual differences in cognitive characteristics underpin different types of responses to reasoning tasks? *Thinking and Reasoning*, *29*(4), 594–642. https://doi.org/10.1080/13546783.2022.2108897
Field, H. (1978). Mental representation. *Erkenntnis*, *13*(1), 9–61. https://doi.org/10.1007/BF00160888
Fischer, E. (2011). *Philosophical delusion and its therapy*. Routledge.
Fischer, E., Allen, K., & Engelhardt, P. E. (2023). Fragmented and conflicted: Folk beliefs about vision. *Synthese*, *201*(3), 84. https://doi.org/10.1007/s11229-023-04066-w
Fischer, E., Allen, K., & Engelhardt, P. E. (2026). Scientific or naïve? Perceptions of Direct and Indirect Realism, and why they matter. *Philosophy and Phenomenological Research*, *112*(1), 102–129. https://doi.org/10.1111/phpr.70063
Fischer, E., & Engelhardt, P. E. (2020). Lingering stereotypes: Salience bias in philosophical argument. *Mind & Language*, *35*(4), 415–439. https://doi.org/10.1111/mila.12249
Fish, W. (2009). *Perception, hallucination, and illusion*. OUP.
Fodor, J. (1987). *Psychosemantics*. MIT Press.
Foisy, L.-M., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education*, *4*(1–2), 26–36. https://doi.org/10.1016/j.tine.2015.03.001
Forstmann, M., & Burgmer, P. (2022). The cartesian folk theater: People conceptualize consciousness as a spatio-temporally localized process in the human brain. *Journal of Experimental Psychology. General*, *151*(4), 781–803. https://doi.org/10.1037/xge0001108
Frankish, K. (2004). *Mind and Supermind*. CUP.
Funkhouser, E. (2017). Beliefs as signals: A new function for belief. *Philosophical Psychology*, *30*(6), 809–831. https://doi.org/10.1080/09515089.2017.1291929
Gelman, S. A., & Noles, N. S. (2011). Domains and naïve theories. *WIREs Cognitive Science*, *2*(5), 490–502. https://doi.org/10.1002/wcs.124
Genone, J. (2016). Recent work on naïve realism. *The American Philosophical Quarterly*, *53*(1), 1–26.
Graziano, M. S. A. (2022). A conceptual framework for consciousness. *Proceedings of the National Academy of Sciences*, *119*(18), e2116933119. https://doi.org/10.1073/pnas.2116933119

Graziano, M. S. A., & Kastner, S. (2011). Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cognitive Neuroscience*, *2*(2), 98–113. https://doi.org/10.1080/17588928.2011.565121

Gross, C. G. (1999). The fire that comes from the eye. *The Neuroscientist*, *5*(1), 58–64. https://doi.org/10.1177/107385849900500108

Guterstam, A., & Graziano, M. S. A. (2020). Visual motion assists in social cognition. *Proceedings of the National Academy of Sciences of the USA*, *117*(50), 32165–32168. https://doi.org/10.1073/pnas.2021325117

Guterstam, A., Kean, H. H., Webb, T. W., Kean, F. S., & Graziano, M. S. A. (2019). Implicit model of other people's visual attention as an invisible, force-carrying beam projecting from the eyes. *Proceedings of the National Academy of Sciences of the USA*, *116*(1), 328–333. https://doi.org/10.1073/pnas.1816581115

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis: Pearson new international edition*. Pearson.

Hampton, J. A. (2006). Concepts as prototypes. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 79–113). Elsevier.

Harris, P., & Gimenez, M. (2005). Children's acceptance of conflicting testimony: The case of death. *Journal of Cognition and Culture*, *5*(1–2), 143–164. https://doi.org/10.1163/1568537054068606

Hawking, S., & Mlodinov, L. (2010). *The Grand design*. Bantam.

Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin and Review*, *22*(6), 1480–1506. https://doi.org/10.3758/s13423-015-0890-8

Hohwy, J. (2013). *The predictive Mind*. OUP.

Hume, D. (1739-40/2011). D.F., & M.J. (Eds.). *A treatise of human nature*. OUP.

Jackman, M. R. (1973). Education and prejudice or education and response-set? *American Sociological Review*, *38*(3), 327–339. https://doi.org/10.2307/2094356

Jackson, F. (1977). *Perception. A representative theory*. CUP.

Kindermann, D., & Onofri, C. (2021). The Fragmented Mind: An Introduction. In C. Borgoni, D. Kindermann, & A. Onofri (Eds.), *The fragmented mind* (pp. 1–34). OUP.

Knobe, J. (2026). Conflicting intuitions. *Ergo*, *13*(5). https://doi.org/10.3998/ergo.9262

Lane, J. D., Zhu, L., Evans, E. M., & Wellman, H. M. (2016). Developing concepts of the mind, body, and afterlife: Exploring the roles of narrative context and culture. *Journal of Cognition and Culture*, *16*(1–2), 50–82. https://doi.org/10.1163/15685373-12342168

Legare, C. H., & Gelman, S. A. (2008). Bewitchment, biology, or both: The co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science*, *32*(4), 607–642. https://doi.org/10.1080/03640210802066766

Leiser, D. (2001). Scattered naïve theories: Why the human mind is isomorphic to internet web. *New Ideas in Psychology*, *19*(3), 175–202. https://doi.org/10.1016/S0732-118X(01)00007-1

Lewis, D. (1997). Naming the colors. *Australasian Journal of Philosophy, 75*(3), 325–342. https://doi.org/10.1080/00048409712347931

Lindberg, D. C. (1976). *Theories of vision from al-Kindi to Kepler*. University of Chicago Press.

Locke, J., ed. Nidditch, P. (1700/1975). *An essay concerning human understanding* (4th ed.). Clarendon.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*(3), 635–694. https://doi.org/10.2466/pr0.1957.3.3.635

Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design: Evidence of a preference for teleological explanations in patients with Alzheimer's disease. *Psychological Science*, *18*(11), 999–1006. https://doi.org/10.1111/j.1467-9280.2007.02015.x

Lyons, J. (2023). Epistemological problems of perception. In E. N. Zalta, & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy*. Summer 2023. https://plato.stanford.edu/archives/sum2023/entries/perception-episprob/

Machery, E. (2009). *Doing without concepts*. OUP.

Macpherson, F. (2013). The philosophy and psychology of hallucination. In F. Macpherson, & D. Platchias (Eds.), *Hallucination: Philosophy and psychology* (pp. 1–38). MIT Press.

Mareschal, D. (2016). The neuroscience of conceptual learning in science and mathematics. *Current Opinion in Behavioral Sciences*, *10*, 114–118. https://doi.org/10.1016/j.cobeha.2016.06.001

Martin, M. G. F. (2002). The transparency of experience. *Mind & Language*, *17*(4), 376–425. https://doi.org/10.1111/1468-0017.00205

Masson, S., Potvin, P., Riopel, M., & Foisy, L.-M. B. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind Brain & Education*, *8*(1), 44–55. https://doi.org/10.1111/mbe.12043

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 299–324). Erlbaum.

McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *9*(4), 636–649. https://doi.org/10.1037/0278-7393.9.4.636

McDowell, J. (2010). Tyler Burge on disjunctivism. *Philosophical Explorations*, *16*(3), 259–279.

McKinsey, M. (2024). Skepticism and content externalism. In E. N. Zalta, & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy* (Spring 2024 Edition). https://plato.stanford.edu/archives/spr2024/entries/skepticism-content-externalism/

McLaughlin, B. P. (2010). The representational vs. The relational view of visual experience. *Royal Institute of Philosophy Supplement*, *67*, 239–262. https://doi.org/10.1017/S135824611000010X

McRae, K., Ferretti, T. R., & Amyote, I. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*(2–3), 137–176. https://doi.org/10.1080/016909697386835

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *The American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Meyer, M., & Knobe, J. (2026). Dimensions of identity-representing belief. *Cognition*, *268*, 106378. https://doi.org/10.1016/j.cognition.2025.106378

Meyering, T. C. (1989). *Historical Roots of Cognitive Science: The rise of a Cognitive theory of perception from antiquity to the Nineteenth Century*. Kluwer.

Musolino, J., Sommer, J., & Hemmer, P. (Eds.) (2022). Individual differences in beliefs. In *The cognitive science of belief: A multidisciplinary approach* (pp. 463–512). CUP.

Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. T. (2014). The free will inventory: Measuring beliefs about agency and responsibility. *Consciousness and Cognition*, *25*, 27–41. https://doi.org/10.1016/j.concog.2014.01.006

O'Connor, T., & Franklin, C. (2022). Free will. In E. N. Zalta, & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/win2022/entries/freewill

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Pautz, A. (2021). *Perception*. Routledge.

Phillips, D. L., & Clancy, K. J. (1972). Some effects of "social desirability" in survey Studies. *The American Journal of Sociology*, *77*(5), 921–940. https://doi.org/10.1086/225231

Piaget, J. (1979). *The Child's conception of the world*, A. & J. Tomlinson. Little, Adams.

Porot, N., & Mandelbaum, E. (2021). The science of belief: A progress report. *WIREs Cognitive Science*, *12*(2), e1539. https://doi.org/10.1002/wcs.1539

Potvin, P., & Cyr, G. (2017). Toward a durable prevalence of scientific conceptions: Tracking the effects of two interfering misconceptions about buoyancy from preschoolers to science teachers. *Journal of Research in Science Teaching*, *54*(9), 1121–1142. https://doi.org/10.1002/tea.21396

Price, H. H. (1932). *Perception* (2nd ed., repr. 1961), Methuen.

Quilty-Dunn, J., & Mandelbaum, E. (2018). Against dispositionalism: Belief in cognitive science. *Philosophical Studies*, *175*(9), 2353–2372. https://doi.org/10.1007/s11098-017-0962-x

Raineri, C. (2021). What's so naïve about naïve realism? *Philosophical Studies*, *178*(11), 3637–3657. https://doi.org/10.1007/s11098-021-01618-z

Reid, T. (1785/1969). *Essays on the intellectual powers of man*. MIT Press.

Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naive physics reasoning: A commitment to substance-based conceptions. *Cognition & Instruction*, *18*(1), 1–34. https://doi.org/10.1207/S1532690XCI1801_01

Rescher, N. (2001). *Paradoxes: Their Roots, range, and resolution*. Open Court.

Roberts, P., Allen, K., & Schmidtke, K. (2018). Reflective intuitions about the causal theory of perception across sensory modalities. *Review of Philosophy and Psychology*, *12*(2), 257–277. https://doi.org/10.1007/s13164-020-00478-6

Robinson, H. (1994). *Perception*. Routledge.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rumelhart, D. E. (1978). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Erlbaum.
Russell, B. (1912/2000). *The problems of Philosophy*. OUP.
Sant'Anna, A., & Dranseika, V. (2024). Does Macbeth see a dagger? An empirical argument for the existence-neutrality of seeing. *Erkenntnis*, *89*(2), 641–664. https://doi.org/10.1007/s10670-022-00549-3
Schwitzgebel, E. (2023). Belief. In E. N. Zalta, & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy* (Spring 2024 Edition), https://plato.stanford.edu/archives/spr2024/entries/belief/
Searle, J. (2015). *Seeing things as they are: A theory of perception*. Oxford University Press.
Shtulman, A. (2017). *Scienceblind. Why our intuitive theories about the world are so often wrong*. Basic Books.
Shtulman, A., & Legare, C. H. (2020). Competing explanations of competing explanations: Accounting for conflict between scientific and folk explanations. *Topics in Cognitive Science*, *12*(4), 1337–1362. https://doi.org/10.1111/tops.12483
Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209–215. https://doi.org/10.1016/j.cognition.2012.04.005
Siegel, S. (2011). *The contents of visual experience*. OUP.
Simms, L. J. (2008). Classical and modern methods of Psychological scale construction. *Social and Personality Psychology Compass*, *2*(1), 414–433. https://doi.org/10.1111/j.1751-9004.2007.00044.x
Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*(2), 189–228. https://doi.org/10.1207/s15516709cog2202_2
Snowdon, P. F. (1981). XI—perception, vision and causation. *Proceedings of the Aristotelian Society*, *81*(1), 175–192. https://doi.org/10.1093/aristotelian/81.1.175
Sommer, J., Musolino, J., & Hemmer, P. (2023). A hobgoblin of large minds: Troubles with consistency in belief. *WIREs Cognitive Science*, *14*(4), e1639. https://doi.org/10.1002/wcs.1639
Sommers, T. (2010). Experimental Philosophy and Free Will. *Philosophy Compass*, *5*(2), 199–212. https://doi.org/10.1111/j.1747-9991.2009.00273.x
Soter, L. K. (2025). A defense of back-end doxastic voluntarism. *Noûs*, *59*(1), 112–139. https://doi.org/10.1111/nous.12501
Spector, P. E. (1992). *Summated rating scale construction: An Introduction*. SAGE.
Šrol, J., & De Neys, W. (2020). Predicting individual differences in conflict detection and bias susceptibility during reasoning. *Thinking and Reasoning*, *27*(1), 38–68. https://doi.org/10.1080/13546783.2019.1708793
Stevens, J. P. (2012). *Applied Multivariate statistics for the social Sciences* (5th ed.). Routledge.
Stricker, J., Vogel, S. E., Schöneburg-Lehnert, S., Krohn, T., & Dögnitz, S., et al. (2021). Interference between naïve and scientific theories occurs in mathematics and is related to mathematical achievement. *Cognition*, *214*, 104789. https://doi.org/10.1016/j.cognition.2021.104789
Strawson, P. F. (1959). *Individuals*. Methuen.
Strawson, P. F. (1979). Perception and its objects. In G. Macdonald (Ed.), *Perception and identity* (pp. 41–60). Palgrave.
Stroud, B. (1984). *The significance of philosophical Scepticism*. OUP.
Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
Thagard, P. (2014). Explanatory identities and conceptual change. *Science & Education*, *23*(7), 1531–1548. https://doi.org/10.1007/s11191-014-9682-1
Thagard, P., & Nisbett, R. E. (1983). Rationality and charity. *Philosophy of Science, 50*, 250–267. https://doi.org/10.1086/289108
Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning & Instruction*, *4*(1), 45–69. https://doi.org/10.1016/0959-4752(94)90018-3
Watson-Jones, R. E., Busch, J. T. A., Harris, P. L., & Legare, C. H. (2017). Does the body survive death? Cultural variation in beliefs about life everlasting. *Cognitive Science*, *41*(S3), 455–476. https://doi.org/10.1111/cogs.12430
Williams, M. (2001). *The problem of knowledge*. OUP.
Winer, G. A., Cottrell, J. E., Gregg, V., Fournier, J. S., & Bica, L. A. (2002). Fundamentally misunderstanding visual perception: Adults' belief in visual emissions. *The American Psychologist*, *57*(6–7), 417–424. https://doi.org/10.1037/0003-066X.57.6-7.417
Winer, G. A., Cottrell, J. E., Karefilaki, K., & Gregg, V. R. (1996). Images, words, and questions: Variables that influence beliefs about vision in children and adults. *Journal of Experimental Child Psychology*, *63*(3), 499–525. https://doi.org/10.1006/jecp.1996.0060

Wiser, M. F., & Carey, S. (1983). When heat and temperature were one. In D. Gentner, & A. L. Stevens (Eds.), *Mental models*. Erlbaum.
Zaitchik, D., & Solomon, G. E. A. (2008). Animist thinking in the elderly and in patients with Alzheimer's disease. *Cognitive Neuropsychology*, *25*(1), 27–37. https://doi.org/10.1080/02643290801904059
Zumbo, B. D. (2023). A dialectic on validity: Explanation-focused and the many ways of being human. *International Journal of Assessment Tools in Education*, *10*(Special Issue), 1–96. https://doi.org/10.21449/ijate.1406304

Authors and Affiliations

Eugen Fischer[1] · Keith Allen[2] · Paul E. Engelhardt[3]

✉ Eugen Fischer
e.fischer@uea.ac.uk

[1] School of Politics, Philosophy, and Area Studies, University of East Anglia, Norwich, UK

[2] Department of Philosophy, University of York, York, UK

[3] School of Psychology, University of East Anglia, Norwich, UK