Trends in
Cognitive Sciences

CellPress

Review

The Perception of Relations

Alon Hafri [1,2,*] and Chaz Firestone [1,2,3,*]

The world contains not only objects and features (red apples, glass bowls, wooden tables), but also relations holding between them (apples contained in bowls, bowls supported by tables). Representations of these relations are often developmentally precocious and linguistically privileged; but how does the mind extract them in the first place? Although relations themselves cast no light onto our eyes, a growing body of work suggests that even very sophisticated relations display key signatures of automatic visual processing. Across physical, eventive, and social domains, relations such as SUPPORT, FIT, CAUSE, CHASE, and even SOCIALLY INTERACT are extracted rapidly, are impossible to ignore, and influence other perceptual processes. Sophisticated and structured relations are not only judged and understood, but also *seen* — revealing surprisingly rich content in visual perception itself.

Highlights

The world is more than a 'bag of objects': it contains relations and interactions between entities. How are such relations extracted by the mind?

Recent work using the tools of vision science suggests that, just as the visual system computes properties such as an object's color or shape, it also computes an object's relation to its physical and social environment, automatically categorizing configurations of objects into distinct relational types.

Across physical, eventive, and social domains, sophisticated and structured relations such as SUPPORT, FIT, CAUSE, and CHASE are not only judged and considered, but also *seen*.

The perception of such sophisticated relations reveals surprisingly rich content in visual perception itself, with consequences for longstanding debates in the philosophy of perception and computational modeling of human vision.

Seeing and Thinking about Relations between Objects

Look at the image in Figure 1A; what do you see in it? Certainly you see some objects (two puzzle pieces), their features (blue, matte, roughly square), and their placement on the page in front of you. However, beyond the individual objects themselves, you may also see something else: the two pieces can fit into one another. This impression captures a relation: a property holding between the objects, beyond any properties each has on its own. What are relations, and how do we represent them?

Relational representations touch nearly all corners of cognitive science, including linguistics (as in relational terms like 'in', 'on', or 'before'), cognitive development (as when children make inferences about interacting objects or agents), analogical reasoning (as when we map entities from one domain to another), and more. However, they are much less emphasized in perception research itself. Leading vision science textbooks devote chapters to motion, color, size, depth, and even objects, faces, and scenes [1,2]; yet they rarely discuss the relational properties you may experience in Figure 1: 'fitting into', 'resting upon', and so on. On one hand, this is understandable: whereas each puzzle piece subtends some visual angle on the retina, there is no component of the retinal image corresponding to their 'fitting'. Indeed, one might suppose that fitting here is not seen at all, but instead only judged, considered, or thought about — much as we might consider whether the puzzle pieces are expensive, appropriate for children, or made by hand.

On the other hand, a recent and growing body of work suggests that we do not only reason about such sophisticated relations in moments of deliberate reflection, but also *see* them directly, much as we see properties like shape, motion, or color. In this review, we discuss several key properties of such relations, and we delineate specific criteria for implicating automatic visual processing (as distinct from higher-level judgment or reasoning). We then apply this framework to empirical work exploring relational perception across several domains, including physical, eventive, and social relations. That such sophisticated relations are properly perceived reveals surprisingly rich content in visual processing and raises new possibilities about the function of perception itself.

[1]Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218, USA
[2]Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218, USA
[3]Department of Philosophy, Johns Hopkins University, Baltimore, MD 21218, USA

*Correspondence:
alon@jhu.edu (A. Hafri) and
chaz@jhu.edu (C. Firestone).

https://doi.org/10.1016/j.tics.2021.01.006

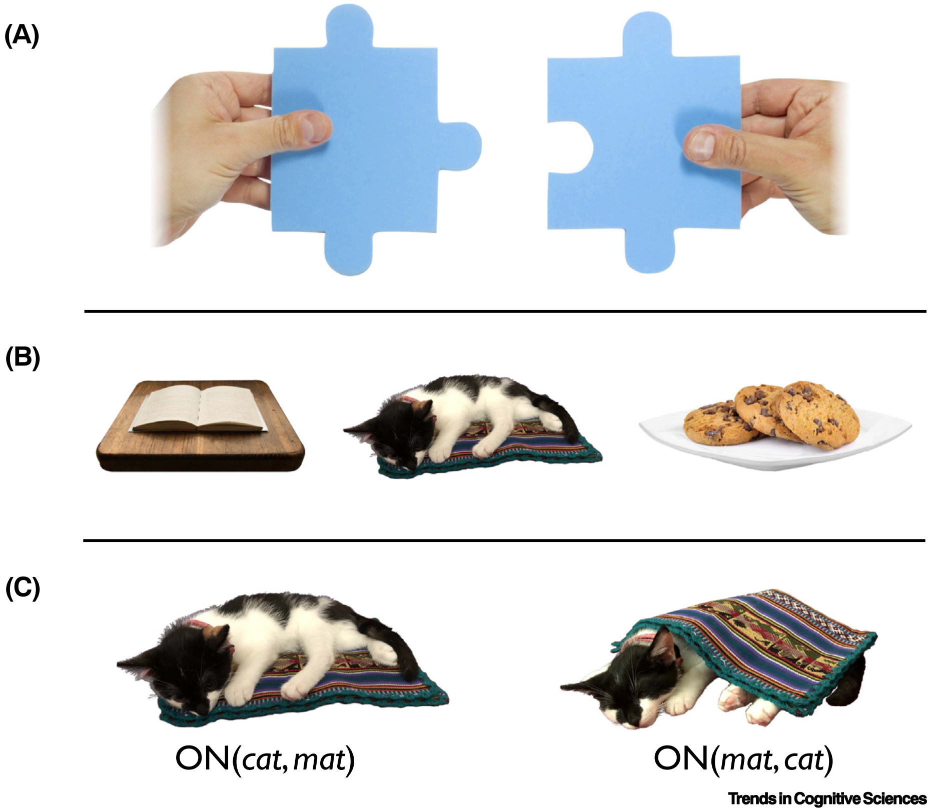

Figure 1. Sophisticated Relations in Natural Images. (A) Two puzzle pieces have visual properties of their own (e.g., colors, shapes, locations), but they also bear a relation to one another: they fit together. (B) Like other relations we consider here, the SUPPORT relation depicted above generalizes beyond any one instance and holds over arbitrary entities. Just as a book on a table is an instance of SUPPORT, so too is a cat on a mat or cookies on a plate — and we can see this commonality even when each instance involves different objects and features. (C) Relations are structured: in other words, 'order' matters. Here, both images contain a cat and a mat and even involve the same relation (SUPPORT); however, in one case a cat is on a mat, while in the other a mat is on a cat. The roles of these entities are distinct and non-interchangeable, and so the structure of these two scenes is different. This distinction is represented in the figure by the reversed order of arguments in the expressions ON(*cat,mat*) versus ON(*mat,cat*).

Representing Relations: Beyond Space and Magnitude

It is not controversial that we perceive relations of some sort. For example, when an object looks to be some distance away (e.g., from us, or from a landmark), or when one stimulus looks bigger, brighter, or bluer than another, they look that way in virtue of relational properties. However, the kinds of relations we explore below — corresponding roughly to what have been called 'functional' or 'force-dynamic' relations [3–5] — differ from comparisons of space or magnitude in that they are surprisingly sophisticated and fundamentally 'interactive'. For example, they may involve the transfer of force between entities, dynamic events that unfold over time, or even a kind of social engagement. Moreover, relations of this sort share several other characteristics that, considered together, distinguish them from other contents represented by our minds.

Relational

Relations require 'relata'. A cat can be on a mat, but it cannot simply be *on*, period; unlike being red or round, it takes two objects to instantiate the relation ON. Though this may seem obvious (and is true by definition), the mind may exploit this fact in representing relational content. For example, someone who hears a sentence beginning ‘Lisa went to…’ knows that the sentence must continue, because a second relatum is required. And even a complete sentence (‘Lisa ate’) can imply an unmentioned relatum: Lisa ate *something* (e.g., her soup). Below, we discuss evidence that visual processing ‘fills in’ relational details in surprisingly similar ways. For example, when a mime tugs an ‘invisible rope’ or bumps into an ‘invisible wall’, our minds automatically supply the implied relatum, actively representing the participating objects.

Abstract

Relations generalize beyond particular instances, object categories, or features. For example, Figure 1B depicts different objects in SUPPORT relations: a book on a table, a cat on a mat, and cookies on a plate. The relation SUPPORT is sufficiently abstract to hold over all such cases — indeed, over any object that could be supported, in arbitrary combination with any object that might support it. Below, we discuss evidence that visual processing embraces this generality: just as two red objects can appear similar while sharing few other features, a book on a table and a cat on a mat are perceived as similar, despite involving very different objects and low-level properties.

Categorical

Whereas metric comparisons are continuous (e.g., one object may be any distance away from another object), the relations considered here are typically ‘all-or-nothing’. For example, one object may be inside another, or it may not be; but there isn’t much middle ground. Below, we discuss evidence for such categorical representation: though there can be better or worse examples of INSIDE, visual processing draws a sharp distinction between INSIDE and OUTSIDE.

Structured

For many relations, ‘order’ matters, such that $R(x,y)$ may be very different from $R(y,x)$. For example, the two images in Figure 1C involve the same objects and relation (cats, mats, and SUPPORT). However, cat-on-mat is a very different scene from mat-on-cat; *ON*(*cat*,*mat*) and *ON*(*mat*,*cat*) map to different scenarios. Though there may be exceptions or special cases (e.g., symmetric relations such as *John and Mary meet*) [6], relations are generally taken to be ‘structured’, with the relata assuming non-interchangeable roles (e.g., Figure versus Ground, Agent versus Patient). Below, we discuss evidence that visual processing respects this structure: relations involving the same relata but different structures are perceived as different.

The capacity to bind arbitrary entities to distinct roles — sometimes called ‘role-filler independence’ [7,8] — makes relational representations especially flexible and powerful (and may explain why they arise in so many cognitive domains [7–13]). It also makes such representations quite unlike those typically associated with visual processing (e.g., perceiving an object’s color, shape, or orientation), including even seemingly sophisticated processes such as visual statistical learning. For example, observers readily extract statistical associations between items [14,15] in ways that enhance processing of objects and their typical features or locations [16–21]. However, such regularities are usually stimulus- or category-specific: learning that toasters appear on kitchen counters says little about which things appear on other things in general (e.g., birds’ nests on tree branches). By contrast, the relations we explore here are general, holding over arbitrary entities rather than particular instances. In other words, even if one has only seen cats on mats, and never mats on cats (Figure 1C), one can immediately see how the elements relate.

Isolating Perception: 'Signatures' of Visual Processing

What we see is different from what we think, infer, judge, or understand. Suppose you are shown a car at the dealership. On one hand, you may apprehend the car's color, shape, or size; these properties are seen. On the other hand, you may apprehend how fuel-efficient the car is, how well it handles bad weather, or how popular it might be; these properties are only judged or inferred. Although your impression of a car's popularity may well be *based on* its visual appearance (e.g., its neon color or sleek silhouette), you use this information to reason about popularity, not to perceive popularity. The question at the heart of this review, then, might be phrased as follows: Are SUPPORT, FIT, CAUSE, CHASE, and so on processed more like color, or more like popularity?

One reason this question can be tricky is that we often see and reason about the very same property. For example, we can directly see that a firetruck is red, but we may also indirectly conclude that it is red even without literally seeing its redness (e.g., if we spot its characteristic shape on a dark night, or hear its siren from a distance). How can we separate these processes?

The approach we take here is to enumerate several telltale 'signatures' that distinguish visual perception itself from higher-level cognitive processes such as reasoning or judgment (see especially [22]). In general, *seeing* exhibits most or all of these signatures, whereas more deliberate reasoning exhibits few or none of them. Moreover, recent work in vision science has developed methodological 'tools' to reveal these signatures, so that they are not just theoretical claims about seeing and thinking but also testable criteria for studying this distinction experimentally (Box 1).

Box 1. Seeing versus Thinking: Tools for Isolating Perception

Visual processing exhibits 'signatures' that distinguish it from higher-level reasoning or judgment. Several methodological tools have emerged to reveal these signatures and are increasingly used to study relational perception (Box 2).

Brief Exposure. Automatic visual processing proceeds faster than more deliberative judgments; and so the ability to extract a property after very brief exposures (especially when masked) suggests a fast visual process. For example, topological relations [33], stability [38,39], and even event roles [65] can be perceived after exposures of < 100 ms (Figure IA). Even stronger evidence for speed arises when response times are also constrained to be rapid (e.g., in continuous image sequences [87]).

Adaptation. Visual adaptation for a property suggests dedicated mechanisms for detecting that property, especially when such adaptation is retinotopic. For example, just as sustained viewing of upward motion biases subsequent motion processing downward, sustained viewing of causal launches biases ambiguous events toward non-causal percepts (Figure IB) [53,56]. This pattern is driven by causality *per se*, as adaptation does not occur for non-causal interactions matched on spatiotemporal factors (e.g., 'slipping').

Continuous Flash Suppression. When the two eyes receive inputs of very different salience (e.g., an ordinary shape in one eye versus a dynamic textured pattern in the other), the more salient stimulus initially dominates, until the suppressed stimulus 'breaks through' awareness several seconds later [107]. Breakthrough speed may indicate the suppressed stimulus's salience, and it cannot be explained by strategic responding (which might reflect higher-level reasoning or judgment [28]), because the participant is not even aware of the suppressed stimulus until breakthrough. This method has shown that causal relations are privileged in visual processing: causal launches enter awareness sooner than non-causal stimuli (Figure IC [51]).

Visual Search. Visually salient stimuli are easier to find in clutter. Recent work shows that popout and search asymmetry arise not only for low-level visual properties (e.g., a red object among green objects) but also higher-level relations, such as social interactions [82] and physical instability (Figure ID [40]).

Altering Other Visual Processes. Perceptual processing can be revealed through influences on other perceptual processes (Figure IE). For example, when two stimuli appear sequentially near one another, we experience 'apparent motion' between them in ways that are immune to explicit knowledge [108]; nevertheless, causal relations can alter such percepts [59,60] in ways that are subjectively appreciable and qualitatively striking (reducing concerns of response-bias and other contaminating factors [28]). Relations may also influence numerosity estimates [109], or even 'warp' perceived space or time [37,49,50,89].

Note that these tools need not correspond one-to-one with the signatures of visual perception reviewed in the main text. For example, visual adaptation alone may be an example of automaticity, sensitivity to subtle visual parameters, and (almost by definition) effects on other visual processes.

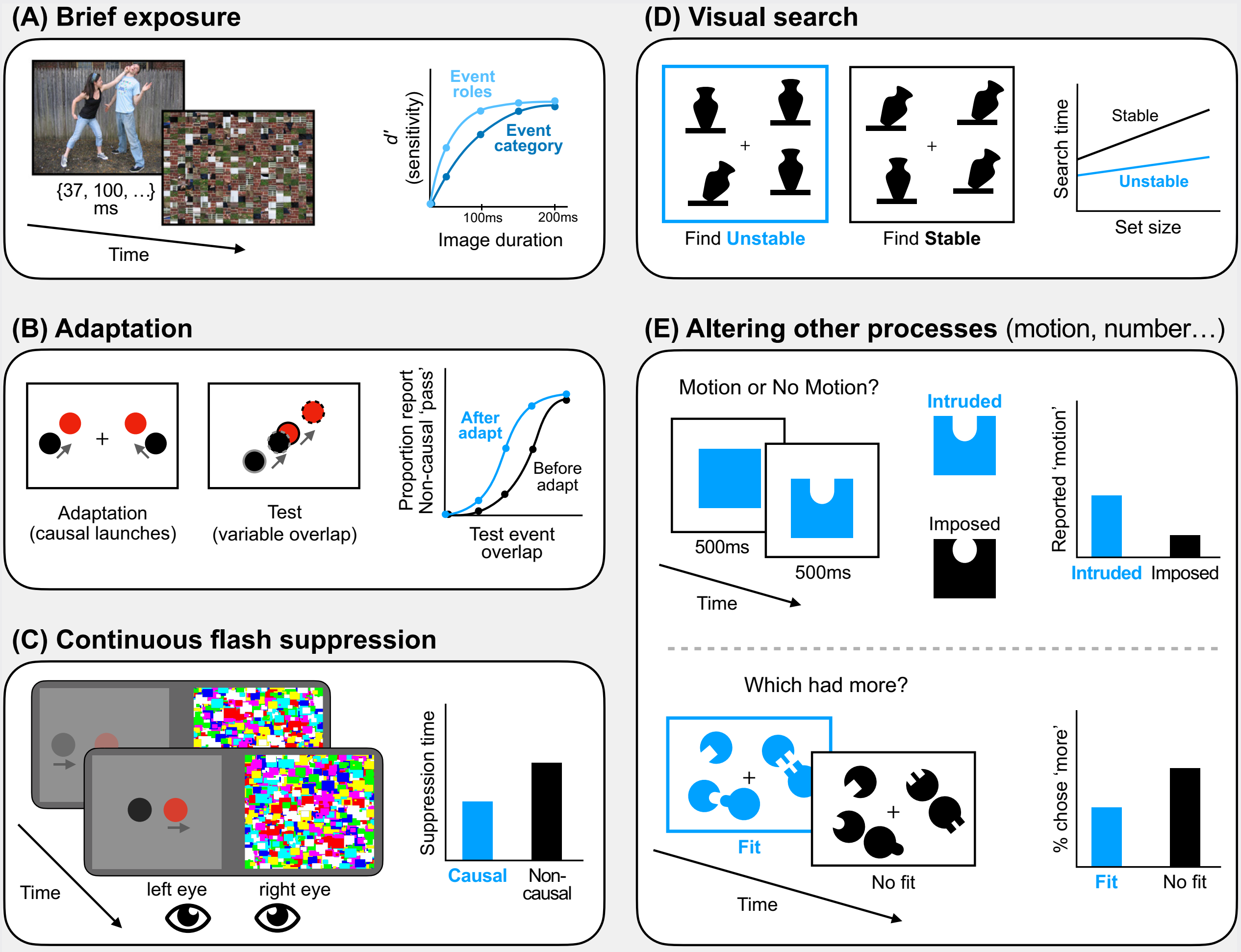

Figure I. Methodological 'Tools' for Implicating Perceptual Processing, as Applied to the Perception of Relations. Many of these paradigms go beyond asking subjects to describe their subjective impressions of the relevant relational properties [149], and instead examine the underlying processing that gives rise to such impressions, or even their effects on other processes. (A) Event roles (who acted upon whom?) can be detected after brief, masked exposures and are processed more efficiently than even the event categories themselves (what action was performed? [65]). (B) Adaptation to causal launches causes ambiguous events to be perceived as non-causal 'passes' [53]. (C) Causal events break through continuous flash suppression faster than non-causal events [51]. (D) Finding an unstable vase among stable vases is easier than finding a stable vase among unstable vases [40]. (E) 'Intruded' shapes generate illusory motion, while 'imposed' shapes do not, an influence of causal history [60]; 'fitting' alters estimates of numerosity [109]. These paradigms often contrast minimal relational pairs whose stimulus properties differ only slightly (e.g., the precise degree of overlap between two shapes) but nevertheless have dramatic perceptual consequences (e.g., whether a causal launch or non-causal pass is perceived). Other paradigms examine specific relational elements (e.g., event roles such as Agent and Patient). Note that many of these phenomena arise not only as subtle effects detectable in controlled settings, but also as rich perceptual experiences that are subjectively striking and phenomenologically apparent. Such phenomenology may even be 'cognitively impenetrable' [28]: even though we know that two circles are not actually causally interacting (since they are simply drawn by a computer program), or that a mime is not really bumping into a wall (since the wall does not even exist), we cannot help but attribute the transfer of force.

Speed

Perception is fast: we can see that something is large, red, or round after extremely brief exposures (tens of milliseconds) and very little processing time (100–200 ms [23,24]). By contrast, determining that a car can safely handle bad weather may require sustained deliberation, even if this deliberation occurs over the car's visual properties. (*Are the treads deep enough? How's the ground clearance?*)

Automaticity

We cannot help but see the world around us: as long as our eyes are open and fixated on a well-lit object, we will perceive that object's color or shape whether we choose to or not. By contrast, one can look at a car *without* reflecting on its safety or fuel-efficiency; it is 'up to us', as it were. One manifestation of such automaticity is that perception often intrudes upon other behaviors. For example, a bright light may capture attention and be impossible to ignore, even when task-irrelevant [25]; but an especially fuel-efficient car doesn't have quite the same effect.

Stubborn Phenomenology

Perception involves not only subtle effects detectable in laboratory settings but also rich experiences that are subjectively striking, even — or especially — when such experiences conflict with more explicit knowledge. For example, we can see an objectively gray object as colored even when we know it isn't (as in the Spanish Castle aftereffect and other color illusions [26]), or see concentric rings as moving even when we know they are static [27]. At most, our explicit knowledge might lead us to disregard or mistrust our visual experiences — but not eliminate their associated phenomenology. Indeed, such stubbornness testifies to the perception/cognition distinction itself: there may be no better way to appreciate how seeing differs from thinking than to perceive the world in a way you know it not to be.

Effects on Other Visual Processes

Whereas perception is stubborn in the face of explicit knowledge, perceptual processes frequently influence one another. For example, an object's perceived distance can alter its perceived size: in the Ponzo illusion, two objects subtending the same visual angle are portrayed as 'close' or 'far' on converging railway tracks; this creates the vivid impression that the objects are different sizes. Such interactions are routine in visual processing; but while higher-level reasoning or judgment may interact with other *cognitive* processes, they rarely influence perception itself (if ever [28]).

Sensitivity to Subtle Visual Parameters

Finally, perception is exquisitely tuned to parameters of visual stimuli that may not otherwise seem notable, such that extremely subtle changes in visual input may dramatically alter the resulting percepts. For example, in the Ternus display [29], two discs appear beside one another in three possible positions (left, center, right), flashing between the left-and-center positions and center-and-right positions. With short flashes, it appears that the central disc is stationary and the other disc is 'jumping over' it (so-called 'element motion'). But if the flash interval is increased by just a few frames, the two discs appear to jump left and right together ('group motion'). The exact point of change — governed by as little as a 10-ms difference — is counterintuitive and nearly impossible to guess in advance; but the perceptual effects it produces are highly reliable and qualitatively striking.

Considered together, these signatures distinguish seeing a visual property from merely judging or reasoning about that property. Of course, other mental processes may exhibit some subset of these signatures (e.g., fast motor reflexes, or even automatic stereotyping); but if a process driven by visual input has most or all of these signatures, it likely reflects *seeing*. Indeed, the very existence of these signatures is itself a reason to pursue this question: evidently, there are systematic generalizations that cleave visual perception from other mental processes [28,30]; and so it is a task for cognitive science to determine where given phenomena lie with respect to this distinction — including the sophisticated relations we explore here.

Structured Relations in Visual Perception: Scope and Evidence

The remainder of this paper explores how automatic visual processing extracts key characteristics of sophisticated relations (involving abstract structure over relata; Box 2). We review evidence across three 'core' domains [31,32]: physical relations (especially static relations, as when one object is in, on, or attached to another), eventive relations (which unfold over time, as when one object pushes, pulls, or deforms another), and social relations (a special type of eventive relation involving interacting agents). Figure 2 illustrates several relations from these domains. Many of these relations have been studied in other areas of cognitive science, including infant cognition and linguistics (Box 3). Do they also arise in visual processing itself?

Physical Relations

Natural scenes are teeming with physical relations: books on shelves, flowers in vases, fences around yards, or apples hanging from trees. Some of these relations are mostly 'spatial' in nature (e.g., the relationship between a fence and the yard it encloses), whereas others imply transmitted or opposing forces, even without any motion or visible change (e.g., books supported by a shelf). How are such relations extracted?

It has long been known that certain spatial and topological relations (e.g., INSIDE versus OUTSIDE) are recognized from extremely brief exposures (< 50 ms [33]). Recent work [34] demonstrates that such processing is automatic and specific to the relational categories themselves, by showing 'categorical perception' for such relations. Participants saw two circles in the relations CONTAINMENT, OVERLAP, TOUCH, or BESIDE and reported whether sequential displays were the same or different. Sometimes the changes also shifted the relational category (e.g., CONTAINMENT to OVERLAP) and sometimes they did not. Discrimination was enhanced for categorical changes, suggesting that relational categories were encoded over and above their metric differences (see also [35,36]). Such processing can even produce visual illusions that 'warp' perceived space depending on whether stimuli appear inside or outside of other objects [37].

Physical relations rapidly alter the deployment of visual attention. An object's stability on a supporting surface (BALANCE) can be determined after masked exposures of only 50–100 ms, even enhancing change-detection for critically unstable blocks [38,39]. Recent work shows that such representations also drive visual search [40]. When search for stimulus A among distractors of type B is faster than search for stimulus B among distractors of type A, this 'search asymmetry' is said to reveal A as a basic visual feature. This work exploited search asymmetry to reveal that physical instability has this property: an unstable object is easier to find among stable objects than a stable object is among unstable objects.

Beyond changing how attention is deployed across different objects, physical relations also modulate perception of the participating objects themselves. Whether an object can be contained depends on its width relative to its would-be container. A recent study [41] suggests that this constraint is wired into the encoding of an object's spatial properties, by finding greater sensitivity

Box 2. Seeing *R*(*x*,*y*): Tools for Isolating Relations

Relational processing involves key characteristics that together distinguish it from other forms of processing (including the extraction of positional regularities or statistical associations [14,15,17–21]). Several tools have emerged to reveal these characteristics and thereby distinguish relations from other contents represented by our minds. Moreover, these tools can be combined with those in Box 1 to isolate relational perception *per se*. For example, when the phenomena below arise through processing that is fast, automatic, and can influence other perceptual processes, this is especially strong evidence for relational perception of the sort we explore here (Figure I).

Categorical Perception (Relational Category). When a category boundary is identified along stimulus dimensions that vary continuously (e.g., distance), this provides evidence that the relational category is represented as such. For example, INSIDE versus OUTSIDE is represented categorically, as shown by improved sensitivity when crossing a category boundary [34] (Figure IA; see also [35,36,52]).

Filling-In (Distinct Relata). Relations — e.g., *R(x,y)* — require both the relation itself (*R*) and the relata (*x* and *y*). When observers mentally fill in the 'missing' item (the elements, or even the relation itself) in fast perceptual tasks, this provides evidence that the relation and its relata are perceived. For example, when an actor bumps into an unseen surface, the mind may impute that surface automatically (Figure IB [66]).

'Confusion' Errors (Abstraction). Relations are abstract, in that they generalize beyond the particular objects involved (e.g., in Figure 1B, the different objects in SUPPORT relations). One tool for exploring this property is to look for 'confusion' errors in speeded recognition tasks. For example, when searching for knife-in-cup (CONTAINMENT), participants false-alarm for other instances of CONTAINMENT (e.g., phone-in-basket), even though those instances involve completely different objects (Figure IC [43]).

Role-Switching (Structure). Relations have structure: John kicking Bill is different from Bill kicking John. One way to probe automatic representation of this structure is to test whether switching it has consequences for visual processing of unrelated features. For example, participants are faster to report an actor's shirt color when that actor maintains the same role across instances (e.g., the Agent) than when that person 'switches roles' from trial to trial (e.g., switches between Agent and Patient) (Figure ID [87]). This demonstrates that the mind encoded the relational structure in parallel with (or even before) the primary task.

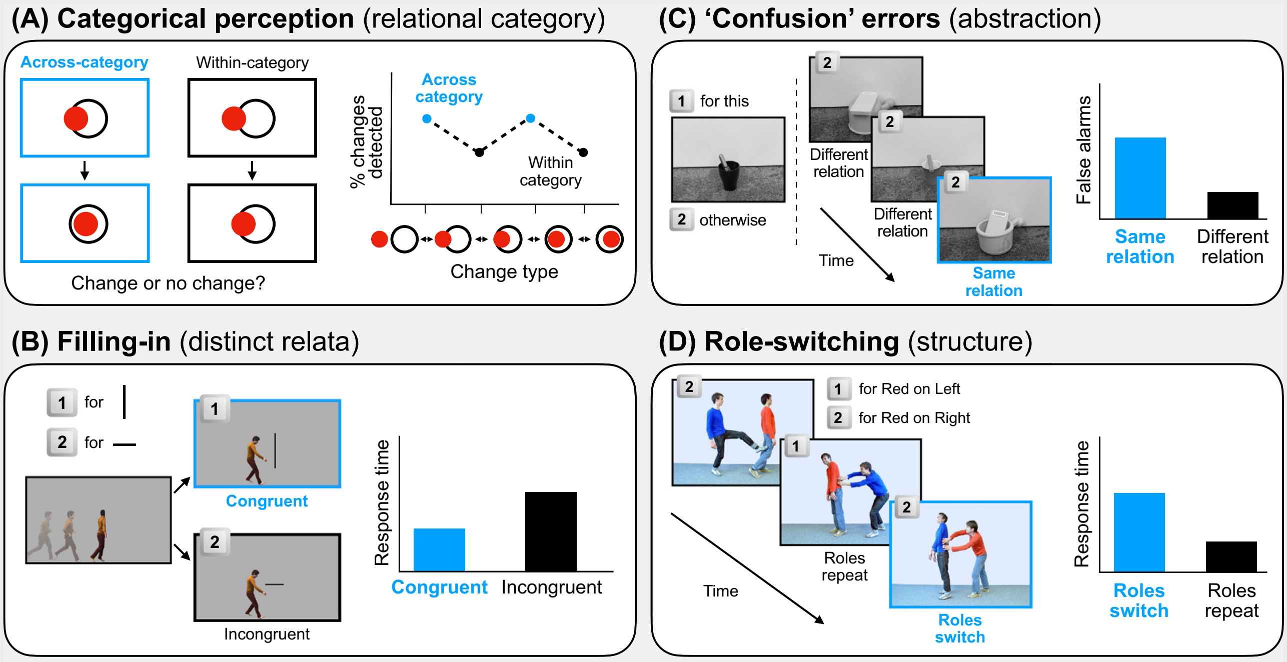

Figure I. Tools for Studying Relations. Just as there are methods to probe perceptual processing *per se* (Box 1), the core characteristics of relations can be assessed in perception studies, as depicted here. (A) Categorical processing of relations (e.g., merely touching versus fully surrounded) can be revealed by increased sensitivity to changes that cross a category boundary [34]). (B) The mind may 'fill in' a relational category or its required relata when those elements do not appear in the display, as when responses to a visible surface are facilitated by first seeing a physical interaction that implies that surface's orientation [66]). (C) Different objects participating in the same relation (e.g., knife-in-cup and phone-in-basket) are encoded as 'similar', as indicated by confusion errors between them [43]). (D) Changes to relational structure (e.g., blue tickling red versus red tickling blue) impair orthogonal perceptual judgments (e.g., who is wearing which color shirt), indicating automatic representation of structured event roles [87]. In general, these paradigms are designed so that the relation itself is incidental or irrelevant to the task given to participants. For example, nothing about reporting the orientation of a line ([66]; B) or the color of an actor's shirt ([87]; D) requires attending to the actors' behavior — indicating that these relational elements were extracted spontaneously or automatically.

to width changes when objects are contained (versus occluded). The authors ruled out several lower-level explanations, such as differences in shading or contour-angles between occluders and containers — suggesting that sensitivity differences were driven by the relational category itself rather than associated lower-level features. 'Fitting' of this sort also interacts with visual recognition: in a recent study [42], participants had to identify a target 'tetromino' (a square composed of Tetris-styled elements) among a stream of distractor tetrominoes. Some distractors could create the target in combination, while others could not. Surprisingly, participants false-alarmed more often to combinable objects (those that could fit together to create the target) than non-combinable ones. In other words, FIT was computed automatically, as indicated by its intrusion on shape recognition.

As noted earlier, a hallmark of relations is their generality, in that very different objects can participate in the same relations; is perception sensitive to this property? A recent study asked participants to identify a target image among a stream of distractors [43]. The images were of different household objects participating in CONTAINMENT or SUPPORT relations (e.g., a phone contained inside a basket, a marker resting on a trashcan, or a knife sitting inside a cup). Intriguingly, participants false-alarmed more often to images that matched the target's relational category than to those that did not — even when such images involved completely different objects. In other words, when searching for a phone in a basket, participants mistakenly responded to a knife in a cup more often than to a marker on a trashcan, suggesting that their minds automatically represented both images as instances of CONTAINMENT.

Eventive Relations

Physical relations acquire a new dimension when they become dynamic: objects may not only passively enclose, support, or contain one another, but also actively push, pull, or deform one another. Such events often have beginnings, endings, and even discrete 'moments' when they occur, and they typically alter the location or state of the objects involved. Are these relations extracted by automatic visual processing?

The study of visual events dates at least to Michotte's investigations of now classic 'launching' stimuli [3,150]. When one disc (A) approaches another (B) and then A stops just as B moves, observers experience the transfer of force from A to B, as if A caused B to move. Though such displays have always been phenomenologically compelling (see discussion in Gibson [44,45]; also [4,46]), recent work has enriched their study by employing the tools of modern vision science and exploring richer and more naturalistic events.

Canonical launching events interact with other processes in visual cognition. For example, causal interaction may cause observers to misreport the location [47] or timing [48,49] of the relevant events; if an event is causal, the distance between causer and causee may be underestimated [50]. Causal relations are even 'privileged' in visual awareness: launches break through continuous flash suppression faster than non-launches do [51], and even very subtle differences between causal events can create categorically different percepts that are salient in visual search arrays [52].

Perhaps the most extraordinary demonstration that causal launching is genuinely *seen* comes from a study [53] exploring a foundational signature of perception: retinotopic visual adaptation (as when, e.g., staring at upward motion makes subsequently viewed stimuli appear to move downward). After viewing multiple launching events, observers were shown test events that varied the discs' spatial overlap; with full overlap, the percept is often of A 'passing by' B. Observers reported whether the test event was a (causal) launch or (non-causal) pass. Remarkably, adapting to causal launches made ambiguous events look like non-causal passes. This effect

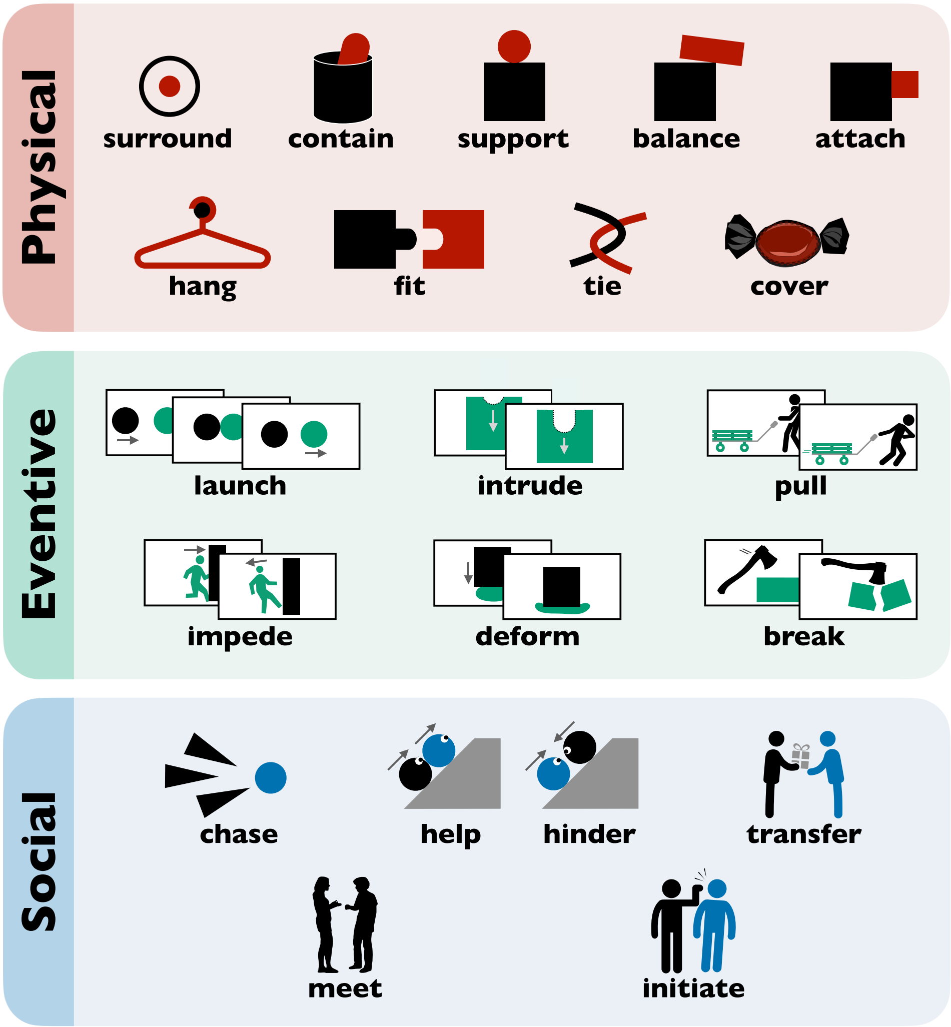

Figure 2. Three Broad Domains of Structured, Categorical Relations in Visual Scenes. Physical relations (red) may imply opposing forces between objects, even without any visible motion or change (as when one object statically supports another). Eventive relations (green) unfold over time and often (but not always) involve discrete 'moments' in which a relation is instantiated (e.g., the point of contact in a causal launch). Social relations (blue) require animate or intentional agents, though the cues to such agency may well arise from simple stimuli that would not otherwise appear animate (as in the above depictions of chasing, helping, or hindering), rather than recognizable human figures. Relations often involve asymmetric structure, with one entity exerting physical or social force on another; here, this is represented by the receiving entity appearing in color. Many of the relations depicted in this figure have been the subject of recent work exploring signatures of perceptual (Box 1) and relational (Box 2) processing and are described in this review. However, some of them (including ATTACH, HANG, TIE, DEFORM, HELP, HINDER, MEET, and TRANSFER) have not yet been investigated for perceptual signatures of relational processing (to our knowledge) and so represent opportunities for future investigation (as do many other relations not depicted here).

did not occur for non-causal 'slipping' events that matched the launches on various low-level properties — and, especially remarkably, was specific to the retinotopic location of the adapting stimulus. No higher cognitive process is known to demonstrate such retinotopic adaptation, making it especially powerful evidence for the perceptual nature of this phenomenon (e.g., as opposed to schema-based theories [54]; see also [55,56]).

Eventive relations can also alter motion processing, including in subjectively appreciable 'demonstrations' that reduce concerns about task demands and other contaminating factors [28,57,58]. When two objects appear sequentially in nearby locations, observers experience 'apparent motion' between them. The inferred motion path is usually the shortest such path; however, causal eventhood can distort this process, producing tortuous motion paths instead of more typical linear trajectories [59]. Causal events can even create motion experiences out of thin air: when a complete shape (e.g., a square) suddenly loses a bite-shaped piece, observers misperceive this discrete change as gradual, falsely reporting the appearance of an intermediate frame as if they had witnessed the 'biting' event itself (an example of 'causal history' [60]; also [61]).

Beyond stripped-down displays with simple stimuli, observers also perceive eventive relations in richer, naturalistic scenes. For example, causal history alters motion perception not only for 'bitten' shapes, but also for actions between human figures (e.g., throwing and catching [62]). Relational properties of such events are also identified rapidly [63–65]: even after masked exposures of only 37 ms (less than a single fixation) to relations such as PUSH, PULL, and KICK, observers immediately extract the event's structure (who was the Agent and who was the Patient), in addition to recognizing its category (what action was performed).

Eventive relations not only show signatures of perceptual processing, but also exemplify hallmarks of relational processing. As noted earlier, dyadic relations take the form *R*(*x*,*y*), where *R*, *x*, and *y* are each necessary components (though other relations may include more elements). Intriguingly, a recent study suggests that the mind 'fills in' the required elements when they do not appear in the display [66]. Actors were filmed interacting with objects (e.g., running into a wall, or stepping onto a box) and then the objects (the wall/box) were digitally removed. These animations (with absent objects) provoked the mind to represent such objects automatically, creating vivid impressions of the 'invisible' wall or box necessary to explain the actor's behavior. These impressions then facilitated responses to actual objects and surfaces appearing moments later, as if the mind was 'primed' by the filled-in objects. In other words, given *R* (e.g., IMPEDE), and *x* (the actor), the mind inferred *y* (the wall). Similar filling-in occurs for other relational elements. For example, if observers are shown only an event's lead-up and consequences (e.g., a foot about to strike a ball, and then the ball soaring in mid-air), they misremember having seen the impact itself (here, perhaps, filling in *R* given subevents involving *x* and *y* [67]; also [68,69]).

Social Relations

Objects and agents interact with each other not just physically, but also socially: one person can push another, but the same person can also help, hinder, pursue, or avoid another. In addition to requiring animate or intentional agents, such relations frequently occur 'at a distance'; two agents need not be in physical contact to have a conversation or chase each other around a playground. In non-social domains, non-contact relations often recruit more effortful processing [9,70–75]; are social relations processed differently?

As with eventive relations, social relations can be perceived from minimal displays that reveal their essential characteristics. In classic work by Heider and Simmel [76], simple geometric shapes moving in self-propelled ways evoke rich narratives involving interacting agents. Recently, similar displays have shown signatures of automatic visual processing.

The perception of chasing appears to be automatic. A recent study [77] created displays in which a pack of moving 'wolves' (dart-shapes) points toward a 'sheep' (an observer-controlled disc), as if pursuing it. Even when given an orthogonal foraging task (such that the darts were task-irrelevant), participants' foraging was impaired when the darts seemed to pursue them, compared to when the darts were oriented perpendicularly or toward another disc. In other words, observers 'couldn't help' but extract the antagonistic relation between wolf and sheep. Impressions of chasing also depend on very subtle display parameters, such as the precise trajectory and orientation of the chasers ('chasing subtlety' [78]). Moreover, chasing is represented as a proper relation, whereby its 'units' are precisely those prioritized in mid-level vision: discrete visual objects [79]. When observers describe displays containing 'wolf' dots and 'sheep' dots, they use mental-state language such as 'evade' and 'follow'; however, when the wolf and sheep are connected to distractors by thin lines, this subtle manipulation breaks the perception of chasing, eliciting far less mental-state language and even making such chasing harder to detect [80]. Evidently, CHASE requires its relata to be discrete objects, not just any visual features (see also [81]).

Beyond geometric shapes, recent work has captured other social relations by including human figures. Social interactions between two people (e.g., arguing, laughing together) show a search asymmetry as described earlier: locating socially interacting individuals among non-interacting individuals is easier than vice versa [82]. Moreover, this effect disappears when the pairs are inverted or facing away from one another, isolating the interaction itself rather than the mere presence of oriented entities, which would not require sophisticated relational processing to extract (see also [83–85]; cf. [86]).

Social relations involving human figures show additional perceptual signatures. Recognition of social interactions can occur in less time than it takes to make a single eye movement [63–65,87,88], and social relations interact with perceptual grouping in ways that alter

Box 3. Where to Look? Insights from Development and Language

We can conceive of a limitless number of relations: not only 'A fits into B' or 'A supports B' (see Figure 1 in the main text), but also arbitrary or contrived relations such as 'A covers exactly 1/3 of B'. Which relations are most likely to show signatures of perceptual processing? Although one's own subjective experience is certainly informative, cognitive development and linguistics have served as surprisingly useful guides to which relational representations may arise in automatic visual processing [32,94].

What Infants Are Prepared to Notice

Infants do not enter the world as 'blank slates'; they are predisposed to represent certain kinds of information, including relational information. For example, infants understand that objects will fall if unsupported, implying that their minds represent SUPPORT. However, there is no evidence that they represent the relation COVER EXACTLY 1/3 OF (versus, say, 1/4 or 1/2). Relations that infants naturally represent have a history of appearing in automatic visual processing, as in CONTAINMENT, SUPPORT, IMPEDE, and CAUSE [110–113]. Other relations privileged in development but not yet investigated for signatures of relational perception include HELP and HINDER [114,115], BREAK [116], and TRANSFER [117].

How Languages 'Package' Information

Most languages have a word for SUPPORT (e.g., English 'on'; Hebrew '*al*'; French 'sur'). By contrast, no known language has a term for COVER EXACTLY 1/3 OF. In other words, although languages can in principle refer to all sorts of *ad hoc* relations, only some relations are 'packaged' systematically [118]. In some cases, such packaging is as straightforward as lexicalization (e.g., 'in', 'on'). In others, packaging arises through restrictions on acceptable use. For example, Korean lexicalizes not only FIT but also FIT TIGHTLY ('*kkita*'; e.g., a key into a lock) and FIT LOOSELY ('*nehta*'; e.g., a chair into a pickup truck [119]). However, English achieves the same distinction by other means. For example, consider the English verb 'insert': common usage permits one to 'insert a key into a lock' (tight), or even 'insert a key into a pencil sharpener' (also tight, albeit unusual), but not quite 'insert a chair into a pickup truck' (which sounds strange and even unacceptable; a pickup truck is a perfectly fine place for a chair, but one doesn't put it there by 'insertion'). Thus, the objects' geometric properties constrain the usage of this relational term (see also [120–122]), whereas other properties (e.g., their colors) do not [118]. Cross-linguistic investigation of this sort has revealed core cognitive distinctions that turn out to arise in perception, such as event roles (e.g., Agent, Patient [123]). Other packaged distinctions not yet tested in vision include 'eventive symmetry' (e.g., 'the truck and car collide' [6]), boundedness (e.g., ENTER versus TRAVERSE [124,125]), and the manner versus result of causal events (e.g., POUR versus FILL [126]).

distance estimates [89] and working-memory capacity [83]. There is even visual adaptation for social relations — in particular, 'social contingencies' such as giving/taking or leading/following [90]. When observers view such an event (e.g., giving) followed by an event that is ambiguous between a matching action (taking) or non-matching action (catching), the ambiguous action is more often recognized as the non-matching action.

Finally, social interactions show a core characteristic of relations: their structure is explicitly represented, such that $R(x,y)$ is distinguished from $R(y,x)$. A recent study [87] showed observers naturalistic photographs of two-person interactions involving asymmetric roles (e.g., biting or tickling, which each require an Agent and Patient), within a rapid, continuous sequence. Participants completed a simple color-search task in which they indicated the location (left/right) of a target individual (e.g., the actor wearing blue). Intriguingly, a 'switching cost' emerged: responses were slower when the target's role switched from trial-to-trial (e.g., the blue-shirted individual switched from Agent to Patient). In other words, observers extracted this interaction's 'structure' (blue tickling red distinct from red tickling blue) and even did so automatically (since event roles were task-irrelevant). Moreover, the roles extracted were genuinely abstract Agent/Patient roles, since the switching cost generalized across relational category (e.g., from biting to tickling).

In summary, across physical, eventive, and social domains, relations between entities show key signatures of automatic visual processing: they are extracted rapidly and automatically, are sensitive to subtle visual parameters, and interact with other perceptual processes. And the resulting representations show central characteristics of sophisticated relations: they are abstract, categorical, and structured, and they operate over distinct relata.

Seeing 'How': Implications of Relational Perception

A venerable tradition in cognitive science defines visual perception as the capacity 'to know what is where by looking' — to represent objects and their features, located somewhere in space [91]. The work reviewed here explores a new dimension of this capacity: not only 'what' and 'where', but also 'how' objects are situated in their physical and social environment.

Although it may seem counterintuitive that perception would represent properties so far removed from the retinal image, invocations of such 'hidden' structure were once seen in a similar light in other domains of fast and automatic processing, such as linguistic representation. For example, we now understand that sentence parsing represents elements and structures that are neither heard nor spoken, as when the imperative statement 'Close the door!' implies an unmentioned subject, or when 'The cat who was bitten by the dog meowed' implies that the cat (not the dog) vocalized in fright [92]. The perspective outlined here suggests that relational structure plays an importantly similar role in how we see the world around us, and raises important questions for future research (see Outstanding Questions).

One exciting aspect of this perspective is its suggestion that visual processing has a broader cognitive 'reach' than is traditionally assumed. For example, researchers working under this approach have suggested that mechanisms of relational perception are active in processes such as word-learning (e.g., determining whether a novel verb refers to causal or non-causal events [93]), 'core knowledge' representation in infancy (e.g., discriminating Agents from Patients [94]), and even moral judgment (e.g., attributing blame in a car accident [95]). In this way, visual perception itself may underwrite surprisingly sophisticated inferences, including those more commonly associated with higher-level cognition (Box 4).

Box 4. From Vision Science to Philosophy and Computation

That visual processing itself extracts sophisticated and structured relations has wide-ranging consequences for our understanding of perception. For example, an active debate in the philosophy of perception asks whether the contents of perception are 'rich' or 'thin'. The thin view holds that perception represents only low-level properties — color, shape, motion, etc. [127]. The rich view holds that perception also represents higher-level properties — including kinds such as DOG or CHAIR, and even more sophisticated contents such as AGENT or CAUSE [128,129]. However, much of this debate has taken place 'from the armchair', as it were, appealing mostly or only to the phenomenology of perception itself. By contrast, the work reviewed here suggests another way to approach this question: harnessing the tools of vision science to ask whether relations such as CAUSE, SUPPORT, FIT, and CHASE show signatures of visual processing. Indeed, it is increasingly popular to discuss such signatures in philosophical treatments of this question [130–133]. And as we have suggested, we think the empirical case favors the 'rich' view in ways that have only recently become evident.

Similarly, this perspective may motivate a reconsideration of the format of perceptual representations, which are generally considered iconic ('picture-like') rather than propositional ('sentence-like') (for discussion, see [134,135]). In particular, the abstract and structured nature of relational representations — including their implementation of role-filler independence [7,8] — makes them difficult to capture with purely iconic formats (where each 'part' of the representation must correspond to some part of the represented scene [135]). Yet, such structure is easily accommodated by propositional formats employing discrete symbols and compositional rules for representing them. These considerations may invite propositional content into perception itself, making better sense of its interface with higher cognitive processes by 'outputting representations that are immediately consumable by cognition' [136].

Finally, this perspective may have implications for modeling visual processing computationally and even for designing artificial intelligence systems that reproduce such processing. Structured, symbolic representation has often been considered necessary for core cognitive processes such as analogical reasoning and linguistic understanding (for discussion, see [12,151]). However, these same elements have been less emphasized for visual perception itself. Indeed, most of today's leading machine-vision systems, including recent 'deep learning' approaches, generally rely on successive convolutional and pooling operations that have no explicit representation of distinct entities participating in structured relations. Although such systems now surpass popular 'human-level' benchmarks for classifying objects, scenes, and text [137,138], they lack the deeper conceptual understanding that humans exhibit [139–142]. The work we review suggests that human-level scene understanding relies on explicit representations of relational structure, and in ways that recommend explicitly wiring this capacity into the inferential machinery of machine-vision systems. In other words, instead of 'classification first, relations later', such models might implement joint-inference on image properties and the structured relations that may have produced them. Indeed, certain modeling approaches are pursuing these directions [143], including 'analysis-by-synthesis' approaches (e.g., for the arrangement of objects in scenes [144]) and approaches that explicitly incorporate relational structure [145,146] and compositionality [147] or operate on symbol-like inputs (e.g., independent object regions) instead of only pixels [148].

Relatedly, this perspective may complement or enrich theories of physical scene understanding. Recent work proposes that cognition is equipped with a mental 'physics engine' that solves physical reasoning problems by simulating the unfolding of entire scenes; physical inferences (e.g., what will move where) are then read off the outcomes of these simulations [96–98]. By contrast, the work reviewed here suggests that rapid, automatic visual processing can actively classify configurations of objects into prespecified relational types (e.g., CONTAINMENT or SUPPORT). Such classifications could 'bypass' the more laborious processing of general-purpose simulation algorithms: for example, if CONTAINMENT is perceived, the mind may automatically infer that the contained object will move with its container, even without actively simulating that outcome [99].

Note that, while we take the work in this review to suggest that relations are genuinely perceived, few of the relations discussed earlier have demonstrated every signature of relational perception. Indeed, many relations simply have not been investigated using the tools described in Boxes 1 and 2 (e.g., HANG, DEFORM, COVER, or BREAK, which have often inspired compelling stimuli [100–103]). Moving forward, inspiration for follow-up work may come from other research domains (e.g., cognitive development and linguistics; Box 3), though we also uphold a role for one's own experience in the world (as in work inspired by mimes and puzzle games [42,66]).

Another agenda item for future work is to elucidate the informational basis and computational mechanisms underlying the perceptual extraction of structured relations. Although many studies of relational perception rule out 'low-level' explanations of their observed effects, collections of low-level features must play *some* role in forming the higher-level representations that eventually result [104]. Recent work has begun to show how such high-level relational representations can arise from lower-level input — for example, how computations of object segmentation and border ownership may underlie recognition of CONTAINMENT and SUPPORT [105]. Such approaches highlight how vision science shares a formidable but exciting challenge with work in other cognitive domains: accounting computationally for the binding of entities and roles [8–12,106] (Box 4).

Concluding Remarks

The world is more than a bag of objects: it contains not only isolated entities and features (red apples, glass bowls) but also relations between them (red apples *in* glass bowls). These relations are rich, abstract, categorical, and structured — and there is growing evidence that they are properly perceived. Vision itself furnishes abstract relational representations, in ways that not only scaffold inferences about scenes but also expand our scientific understanding of relational processing into new areas of the mind.

Acknowledgements

For helpful discussion and/or comments on previous drafts, the authors thank Simon Brown, Nicolò Cesana-Arlotti, Ernie Davis, Steve Franconeri, Lila Gleitman, Victor Gomes, Steven Gross, Zoe Jenkin, Barbara Landau, Jorge Morales, Barbara Partee, Ian Phillips, Jake Quilty-Dunn, John Trueswell, Jeremy Wolfe, and members of the Johns Hopkins University Perception and Mind Laboratory. This work was supported by NSF BCS-2021053 awarded to C.F.

Declaration of Interests

No interests are declared.

Outstanding Questions

What are the temporal and spatial constraints on relational perception? Perception is 'fast', but many events unfold over long stretches of time and space. For example, a jack may raise a heavy load so slowly that its motion is undetectable; in addition to perceiving the physical relation SUPPORT, do we also perceive the eventive relation LIFT?

Can new, and even arbitrary, relations come to be represented like more canonical perceptual relations? The mind categorically encodes CONTAINMENT, but not COVER EXACTLY 1/3 OF. However, relations that start as arbitrary can acquire significance through learning. For example, athletes may be sensitive to certain metric relations (e.g., '50 yards from') if the relations have consequences in their sport (e.g., 'within field-goal range'). Do such properties show signatures of relational perception?

Do relations in other sensory modalities exhibit the signatures explored here? We can feel that a purse is full of change, or hear that a glass has been shattered to pieces. Are such relations represented automatically and categorically? Do they involve the binding of arbitrary entities to roles?

Beyond the core domains, how are perceived relations organized in the mind? Do relations in different domains share common coding mechanisms? For example, visual adaptation generalizes between 'triggering' and 'launching'; but does seeing one disc launch another transfer to social relations such as SHOVE, or even STARTLE (a kind of 'social triggering')?

How does the mind represent 'embedded' relations? For example, a cat may be on a mat that is in a box. Is this situation represented as multiple independent relations (with cat-on-mat separate from mat-in-box)? Or as a hierarchically structured relation (with the cat-on-mat SUPPORT relation serving as a relatum for CONTAINMENT)?

References

1. Goldstein, E.B. and Brockmole, J. (2016) *Sensation and Perception*, Cengage Learning
2. Wolfe, J.M. *et al.* (2017) *Sensation & Perception* (fifth edition), Oxford University Press, New York, NY
3. Michotte, A. (1946) *La Perception de la Causalité. (Etudes Psychol.), Vol. VI [Perception of Causality.]*, Inst. Sup. De Philosophie, Oxford, England
4. Scholl, B.J. and Tremoulet, P.D. (2000) Perceptual causality and animacy. *Trends Cogn. Sci.* 4, 299–309
5. Talmy, L. (1988) Force dynamics in language and cognition. *Cogn. Sci.* 12, 49–100
6. Gleitman, L.R. *et al.* (1996) Similar, and similar concepts. *Cognition* 58, 321–376
7. Frankland, S.M. and Greene, J.D. (2020) Concepts and compositionality: in search of the brain's language of thought. *Annu. Rev. Psychol.* 71, 273–303
8. Hummel, J.E. and Holyoak, K.J. (2003) A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.* 110, 220
9. Franconeri, S.L. *et al.* (2012) Flexible visual processing of spatial relationships. *Cognition* 122, 210–227
10. Hummel, J.E. and Biederman, I. (1992) Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* 99, 480–517
11. Jackendoff, R. (2002) *Foundations of Language: Brain, Meaning, Grammar, Evolution*, Oxford University Press
12. Marcus, G. (2001) *The Algebraic Mind*, MIT Press, Cambridge, MA
13. Markman, A.B. and Gentner, D. (1993) Structural alignment during similarity comparisons. *Cogn. Psychol.* 25, 431–467
14. Fiser, J. and Aslin, R.N. (2005) Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 134, 521–537
15. Schapiro, A.C. *et al.* (2013) Neural representations of events arise from temporal community structure. *Nat. Neurosci.* 16, 486–492
16. Bonner, M.F. and Epstein, R.A. (2020) Object representations in the human brain reflect the co-occurrence statistics of vision and language. *bioRxiv* Published online March 11, 2020. https://doi.org/10.1101/2020.03.09.984625
17. Kaiser, D. *et al.* (2014) Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc. Natl. Acad. Sci.* 111, 11217–11222
18. Kaiser, D. *et al.* (2019) Object vision in a structured world. *Trends Cogn. Sci.* 23, 672–685
19. Kaiser, D. *et al.* (2015) Real-world spatial regularities affect visual working memory for objects. *Psychon. Bull. Rev.* 22, 1784–1790
20. Kim, J.G. and Biederman, I. (2011) Where do objects become scenes? *Cereb. Cortex* 21, 1738–1746
21. Võ, M.L.-H. *et al.* (2019) Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Curr. Opin. Psychol.* 29, 205–210
22. Scholl, B.J. and Gao, T. (2013) Perceiving animacy and intentionality: visual processing or higher-level judgment? In *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention* (Rutherford, M.D. and Kuhlmeier, V.A., eds), pp. 197–230, MIT Press, Cambridge, MA
23. Mack, M.L. and Palmeri, T.J. (2015) The dynamics of categorization: unraveling rapid categorization. *J. Exp. Psychol. Gen.* 144, 551–569
24. Thorpe, S. *et al.* (1996) Speed of processing in the human visual system. *Nature* 381, 520–522
25. Yantis, S. (1993) Stimulus-driven attentional capture. *Curr. Dir. Psychol. Sci.* 2, 156–161
26. Daw, N.W. (1962) Why after-images are not seen in normal circumstances. *Nature* 196, 1143–1145
27. Pinna, B. and Brelstaff, G.J. (2000) A new visual illusion of relative motion. *Vis. Res.* 40, 2091–2096

28. Firestone, C. and Scholl, B.J. (2016) Cognition does not affect perception: evaluating the evidence for "top-down" effects. *Behav. Brain Sci.* 39, e229
29. Ternus, J. (1926) Experimentelle untersuchungen über phänomenale Identität. *Psychol. Forsch.* 7, 81–136
30. Pylyshyn, Z. (1999) Is vision continuous with cognition?: the case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–365
31. Carey, S. (2009) *The Origin of Concepts*, Oxford University Press
32. Spelke, E.S. and Kinzler, K.D. (2007) Core knowledge. *Dev. Sci.* 10, 89–96
33. Chen, L. (1982) Topological structure in visual perception. *Science* 218, 699–700
34. Lovett, A. and Franconeri, S.L. (2017) Topological relations between objects are categorically coded. *Psychol. Sci.* 28, 1408–1418
35. Kim, J.G. and Biederman, I. (2012) Greater sensitivity to nonaccidental than metric changes in the relations between simple shapes in the lateral occipital cortex. *NeuroImage* 63, 1818–1826
36. Kranjec, A. *et al.* (2014) Categorical biases in perceiving spatial relations. *PLoS One* 9, e98604
37. Vickery, T.J. and Chun, M.M. (2010) Object-based warping: an illusory distortion of space within objects. *Psychol. Sci.* 21, 1759–1764
38. Firestone, C. and Scholl, B. (2016) Seeing stability: intuitive physics automatically guides selective attention. *J. Vis.* 16, 689
39. Firestone, C. and Scholl, B. (2017) Seeing physics in the blink of an eye. *J. Vis.* 17, 203
40. Yang, Y.-H. and Wolfe, J.M. (2020) Is apparent instability a guiding feature in visual search? *Vis. Cogn.* 28, 218–238
41. Strickland, B. and Scholl, B.J. (2015) Visual perception involves event-type representations: the case of containment versus occlusion. *J. Exp. Psychol. Gen.* 144, 570–580
42. Guan, C. and Firestone, C. (2020) Seeing what's possible: disconnected visual parts are confused for their potential wholes. *J. Exp. Psychol. Gen.* 149, 590–598
43. Hafri, A. *et al.* (2020) A phone in a basket looks like a knife in a cup: the perception of abstract relations. *PsyArXiv* Published online May 5, 2020. https://psyarxiv.com/jx4yg/
44. Gibson, J.J. (1967) In *A History of Psychology in Autobiography* (Vol V) (Boring, E.G. and Lindzey, G., eds), pp. 125–143, Appleton-Century-Crofts, East Norwalk
45. Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston
46. Zacks, J.M. and Tversky, B. (2001) Event structure in perception and conception. *Psychol. Bull.* 127, 3–21
47. Scholl, B.J. and Nakayama, K. (2004) Illusory causal crescents: misperceived spatial relations due to perceived causality. *Perception* 33, 455–469
48. Bechlivanidis, C. and Lagnado, D.A. (2016) Time reordered: causal perception guides the interpretation of temporal order. *Cognition* 146, 58–66
49. Buehner, M.J. and Humphreys, G.R. (2009) Causal binding of actions to their effects. *Psychol. Sci.* 20, 1221–1228
50. Buehner, M.J. and Humphreys, G.R. (2010) Causal contraction: spatial binding in the perception of collision events. *Psychol. Sci.* 21, 44–48
51. Moors, P. *et al.* (2017) Causal events enter awareness faster than non-causal events. *PeerJ* 5, e2932
52. Kominsky, J.F. *et al.* (2017) Categories and constraints in causal perception. *Psychol. Sci.* 28, 1649–1662
53. Rolfs, M. *et al.* (2013) Visual adaptation of the perception of causality. *Curr. Biol.* 23, 250–254
54. Rips, L.J. (2011) Causation from perception. *Perspect. Psychol. Sci.* 6, 77–97
55. Arnold, D.H. *et al.* (2015) An object-centered aftereffect of a latent material property. *J. Vis.* 15, 4
56. Kominsky, J.F. and Scholl, B.J. (2020) Retinotopic adaptation reveals distinct categories of causal perception. *Cognition* 203, 104339
57. Firestone, C. and Scholl, B.J. (2015) When do ratings implicate perception versus judgment? The "overgeneralization test" for top-down effects. *Vis. Cogn.* 23, 1217–1226
58. Valenti, J.J. and Firestone, C. (2019) Finding the "odd one out": memory color effects and the logic of appearance. *Cognition* 191, 103934
59. Kim, S.-H. *et al.* (2013) Perceived causality can alter the perceived trajectory of apparent motion. *Psychol. Sci.* 24, 575–582
60. Chen, Y.-C. and Scholl, B.J. (2016) The perception of history: seeing causal history in static shapes induces illusory motion perception. *Psychol. Sci.* 27, 923–930
61. Spröte, P. *et al.* (2016) Visual perception of shape altered by inferred causal history. *Sci. Rep.* 6, 36245
62. Peng, Y. *et al.* (2020) Causal actions enhance perception of continuous body movements. *Cognition* 194, 104060
63. Dobel, C. *et al.* (2007) Describing scenes hardly seen. *Acta Psychol.* 125, 129–143
64. Glanemann, R. *et al.* (2016) Rapid apprehension of the coherence of action scenes. *Psychon. Bull. Rev.* 23, 1566–1575
65. Hafri, A. *et al.* (2013) Getting the gist of events: recognition of two-participant actions from brief displays. *J. Exp. Psychol. Gen.* 142, 880–905
66. Little, P.C. and Firestone, C. (2021) Physically implied surfaces. *Psychol. Sci.* https://doi.org/10.1177/0956797620939942
67. Strickland, B. and Keil, F. (2011) Event completion: event based inferences distort memory in a matter of seconds. *Cognition* 121, 409–415
68. Bae, G.Y. and Flombaum, J.I. (2011) Amodal causal capture in the tunnel effect. *Perception* 40, 74–90
69. Falck, A. *et al.* (2020) Core cognition in adult vision: a surprising discrepancy between the principles of object continuity and solidity. *J. Exp. Psychol. Gen.* 149, 2250–2263
70. Kluth, T. *et al.* (2019) Does direction matter? Linguistic asymmetries reflected in visual attention. *Cognition* 185, 91–120
71. Logan, G.D. (1995) Linguistic and conceptual control of visual spatial attention. *Cogn. Psychol.* 28, 103–174
72. Roth, J.C. and Franconeri, S.L. (2012) Asymmetric coding of categorical spatial relations in both language and vision. *Front. Psychol.* 3
73. Ullman, S. (1984) Visual routines. *Cognition* 18, 97–159
74. Ullman, S. (1996) Visual cognition and visual routines. In *High-Level Vision: Object Recognition and Visual Cognition*, pp. 263–315, MIT Press, Cambridge, Mass
75. Yuan, L. *et al.* (2016) Are categorical spatial relations encoded by shifting visual attention between objects? *PLoS One* 11, e0163141
76. Heider, F. and Simmel, M. (1944) An experimental study of apparent behavior. *Am. J. Psychol.* 57, 243–259
77. van Buren, B. *et al.* (2016) The automaticity of perceiving animacy: goal-directed motion in simple shapes influences visuomotor behavior even when task-irrelevant. *Psychon. Bull. Rev.* 23, 797–802
78. Gao, T. *et al.* (2009) The psychophysics of chasing: a case study in the perception of animacy. *Cogn. Psychol.* 59, 154–179
79. Scholl, B.J. *et al.* (2001) What is a visual object? Evidence from target merging in multiple object tracking. *Cognition* 80, 159–177
80. van Buren, B. *et al.* (2017) What are the underlying units of perceived animacy? Chasing detection is intrinsically object-based. *Psychon. Bull. Rev.* 24, 1604–1610
81. Wick, F.A. *et al.* (2019) Perception in dynamic scenes: what is your Heider capacity? *J. Exp. Psychol. Gen.* 148, 252–271
82. Papeo, L. *et al.* (2019) Visual search for people among people. *Psychol. Sci.* 30, 1483–1496
83. Ding, X. *et al.* (2017) Two equals one: two human actions during social interaction are grouped as one unit in working memory. *Psychol. Sci.* 28, 1311–1320
84. Papeo, L. (2020) Twos in human visual perception. *Cortex* 132, 473–478
85. Papeo, L. *et al.* (2017) The two-body inversion effect. *Psychol. Sci.* 28, 369–379
86. Vestner, T. *et al.* (2020) Why are social interactions found quickly in visual search tasks? *Cognition* 200, 104270

87. Hafri, A. *et al.* (2018) Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition* 175, 36–52
88. Isik, L. *et al.* (2020) The speed of human social interaction perception. *NeuroImage* 215, 116844
89. Vestner, T. *et al.* (2019) Bound together: social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *J. Exp. Psychol. Gen.* 148, 1251–1268
90. Fedorov, L.A. *et al.* (2018) Adaptation aftereffects reveal representations for encoding of contingent social actions. *Proc. Natl. Acad. Sci.* 115, 7515–7520
91. Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Henry Holt and Co., Inc., New York, NY
92. Jackendoff, R. (1987) On Beyond Zebra: The relation of linguistic and visual information. *Cognition* 26, 89–114
93. Kline, M. *et al.* (2017) Linking language and events: spatiotemporal cues drive children's expectations about the meanings of novel transitive verbs. *Lang. Learn. Dev.* 13, 1–23
94. Strickland, B. (2017) Language reflects "core" cognition: a new theory about the origin of cross-linguistic regularities. *Cogn. Sci.* 41, 70–101
95. De Freitas, J. and Alvarez, G.A. (2018) Your visual system provides all the information you need to make moral judgments about generic visual events. *Cognition* 178, 133–146
96. Battaglia, P.W. *et al.* (2013) Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci.* 110, 18327–18332
97. Kubricht, J.R. *et al.* (2017) Intuitive physics: current research and controversies. *Trends Cogn. Sci.* 21, 749–759
98. Ullman, T.D. *et al.* (2017) Mind games: game engines as an architecture for intuitive physics. *Trends Cogn. Sci.* 21, 649–665
99. Davis, E. and Marcus, G. (2015) The scope and limits of simulation in cognitive models. *arXiv* Published online June 16, 2015. https://arxiv.org/abs/1506.04956
100. Phillips, F. and Fleming, R.W. (2020) The veiled virgin illustrates visual segmentation of shape by cause. *Proc. Natl. Acad. Sci.* 117, 11735–11743
101. Spröte, P. and Fleming, R.W. (2016) Bent out of shape: the visual inference of non-rigid shape transformations applied to objects. *Vis. Res.* 126, 330–346
102. White, P.A. and Milne, A. (1999) Impressions of enforced disintegration and bursting in the visual perception of collision events. *J. Exp. Psychol. Gen.* 128, 499–516
103. Yildirim, I. *et al.* (2016) Perceiving fully occluded via physical simulation. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*
104. Halberda, J. (2019) Perceptual input is not conceptual content. *Trends Cogn. Sci.* 23, 636–638
105. Ullman, S. *et al.* (2019) A model for discovering 'containment' relations. *Cognition* 183, 67–81
106. Yuan, L. *et al.* (2020) Learning the generative principles of a symbol system from limited examples. *Cognition* 200, 104243
107. Tsuchiya, N. and Koch, C. (2005) Continuous flash suppression reduces negative afterimages. *Nat. Neurosci.* 8, 1096–1101
108. Wright, R.D. and Dawson, M.R.W. (1994) To what extent do beliefs affect apparent motion? *Philos. Psychol.* 7, 471–491
109. Guan, C. *et al.* (2020) Possible objects count: perceived numerosity is altered by representations of possibility. *J. Vis.* 20, 847
110. Hespos, S.J. and Spelke, E.S. (2004) Conceptual precursors to language. *Nature* 430, 453–456
111. Leslie, A.M. and Keeble, S. (1987) Do six-month-old infants perceive causality? *Cognition* 25, 265–288
112. Spelke, E.S. *et al.* (1992) Origins of knowledge. *Psychol. Rev.* 99, 605–632
113. Wang, S.-h. *et al.* (2016) Young infants view physically possible support events as unexpected: new evidence for rule learning. *Cognition* 157, 100–105
114. Hamlin, J.K. *et al.* (2007) Social evaluation by preverbal infants. *Nature* 450, 557–559
115. Kuhlmeier, V. *et al.* (2003) Attribution of dispositional states by 12-month-olds. *Psychol. Sci.* 14, 402–408
116. Muentener, P. and Carey, S. (2010) Infants' causal representations of state change events. *Cogn. Psychol.* 61, 63–86
117. Tatone, D. *et al.* (2015) Giving and taking: representational building blocks of active resource-transfer events in human infants. *Cognition* 137, 47–62
118. Talmy, L. (1985) Lexicalization patterns: semantic structure in lexical forms. In *Language Typology and Syntactic Description Volume III: Grammatical Categories and the Lexicon* (Shopen, T., ed.), pp. 57–149, Cambridge University Press, Cambridge
119. Bowerman, M. and Choi, S. (2003) Space under construction: language-specific spatial categorization in first language acquisition. In *Language in Mind: Advances in the Study of Language and Cognition* (Gentner, D. and Goldin-Meadow, S., eds), pp. 387–428, MIT Press
120. Landau, B. (2017) Update on "what" and "where" in spatial language: A new division of labor for spatial terms. *Cogn. Sci.* 41, 321–350
121. Landau, B. (2018) Learning simple spatial terms: Core and more. *Top. Cogn. Sci.* 12, 91–114
122. Landau, B. and Jackendoff, R. (1993) "What" and "where" in spatial language and spatial cognition. *Behav. Brain Sci.* 16, 217–238
123. Dowty, D. (1991) Thematic proto-roles and argument selection. *Language* 67, 547
124. Ji, Y. and Papafragou, A. (2020) Is there an end in sight? Viewers' sensitivity to abstract event structure. *Cognition* 197, 104197
125. Strickland, B. *et al.* (2015) Event representations constrain the structure of language: sign language as a window into universally accessible linguistic biases. *Proc. Natl. Acad. Sci.* 112, 5968–5973
126. Gropen, J. *et al.* (1991) Syntax and semantics in the acquisition of locative verbs. *J. Child Lang.* 18, 115–151
127. Tye, M. (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*, MIT Press, Cambridge, MA
128. Bayne, T. (2009) Perception and the reach of phenomenal content. *Philos. Q.* 59, 385–404
129. Siegel, S. (2010) *The Contents of Visual Experience*, Oxford University Press, Oxford
130. Block, N. (2014) Seeing-as in the light of vision science. *Philos. Phenomenol. Res.* 89, 560–572
131. Helton, G. (2016) Recent issues in high-level perception. *Philos. Compass* 11, 851–862
132. Siegel, S. and Byrne, A. (2016) Rich or thin? In *Current Controversies in Philosophy of Perception* (Nanay, B., ed.), pp. 59–80, Routledge, New York
133. Westfall, M. (2020) Other minds are neither seen nor inferred. *Synthese* Published online September 6, 2020. https://doi.org/10.1007/s11229-020-02844-4
134. Burge, T. (2010) *Origins of Objectivity*, Oxford University Press
135. Kosslyn, S.M. *et al.* (2006) *The Case for Mental Imagery*, Oxford University Press
136. Quilty-Dunn, J. (2020) Concepts and predication from perception to cognition. *Philos. Issues* 30, 273–292
137. Krizhevsky, A. *et al.* (2012) Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Process.* 25, 1106–1114
138. LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
139. Firestone, C. (2020) Performance vs. competence in human–machine comparisons. *Proc. Natl. Acad. Sci.* 117, 26562–26571
140. Lake, B.M. *et al.* (2017) Building machines that learn and think like people. *Behav. Brain Sci.* 40, e253
141. Marcus, G. (2018) Deep learning: a critical appraisal. *arXiv* Published online January 2, 2018. https://arxiv.org/abs/1801.00631
142. Yuille, A.L. and Liu, C. (2020) Deep nets: what have they ever done for vision? *Int. J. Comput. Vis.* Published online November 27, 2020. https://doi.org/10.1007/s11263-020-01405-z
143. Zhu, Y. *et al.* (2020) Dark, beyond deep: a paradigm shift to cognitive AI with humanlike common sense. *arXiv* Published online April 20, 2020. https://arxiv.org/abs/2004.09044

144. Ali Eslami, S.M. *et al.* (2018) Neural scene representation and rendering. *Science* 360, 1204–1210
145. Battaglia, P.W. *et al.* (2018) Relational inductive biases, deep learning, and graph networks. *arXiv* Published online June 4, 2018. https://arxiv.org/abs/1806.01261
146. Bear, D.M. *et al.* (2020) Learning physical graph representations from visual scenes. *arXiv* Published online June 22, 2020. https://arxiv.org/abs/2006.12373
147. Wang, J. *et al.* (2017) Visual concepts and compositional voting. *arXiv* Published online November 13, 2017. https://arxiv.org/abs/1711.04451
148. Kim, J. *et al.* (2018) Not-So-CLEVR: learning same–different relations strains feedforward neural networks. *Interface Focus* 8, 20180011
149. White, P.A. and Milne, A. (1997) Phenomenal causality: impressions of pulling in the visual perception of objects in motion. *Am. J. Psychol.* 110, 573–602
150. Wagemans, J. *et al.* (2006) Introduction to Michotte's heritage in perception and cognition research. *Acta Psychol.* 123, 1–19
151. Holyoak, K.J. and Lu, H. (2021) Emergence of relational reasoning. *Curr. Opin. Behav. Sci.* 37, 118–124