Sensory Substitution and Perceptual Emergence

Jonathan Cohen*

The problems with which we are confronted in our strivings for a scientific understanding of mental phenomena are gigantic, and there is little reason to be overenthusiastic about the present state of the field. But one certitude remains: The mental can be driven out of psychology as little as the biological can be driven out of biology or the chemical out of chemistry. Where no psychology is put in, no psychology is put out (Mausfeld, 2003, 190).

Abstract

Sensory substitution devices (SSDs) can be thought of as sensory prosthetics — as devices by which information normally represented by one perceptual modality is instead represented by an alternative, non-canonical channel involving a second perceptual modality. Given that sensory modalities ordinarily represent a very wide range of properties/events, it is no trivial matter to ensure that such prosthetic devices will capture the representational scope of the systems they aim to replace. Designers of such devices have often hoped to solve this challenge by building SSDs that succeed in representing the form(s) of basic energy normally represented by the impaired modality they are intended to replace, and then adding computational power to code up whatever can be derived from the basic energy. Thus, for example, the thought is that if we could build a device that represents in some alternative way the distal distribution of light intensity — the basic form of energy to which visual receptors are normally responsive — our device could, in principle, represent everything vision can represent: color, shape, form, motion, and so on.

Unfortunately, I will argue, this attractively simple idea fails. For there appear to be features represented by our sensory modalities that are "perceptually emergent" — i.e., features whose exemplification is not fixed by the representation of the distribution of (the relevant forms of) basic energy. Hence, a SSD whose basic representational vocabulary is limited to the distribution of such basic energy will leave things out.

None of this shows that SSDs will inevitably fail to represent what sensory modalities normally represent. It does suggest, however, that if we want them to represent what sensory modalities normally represent, then we will have to do more than preserve the representation of basic energy to which the substituted modalities are sensitive.

*Department of Philosophy, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0119, joncohen@aardvark.ucsd.edu

Keywords — Sensory substitution, perceptual emergence, basic energy, perceptual modalities, endogenous contributions to perception.

Sensory substitution devices (SSDs) can be thought of as sensory prosthetics — as devices by which information represented by one perceptual modality (the substituted modality) is instead represented by an alternative, non-canonical representational channel involving a second perceptual modality (the substituting modality). The hope is that such devices might aid subjects with an impaired perceptual modality in exploiting their unimpaired modalities to connect perceptually with the world in ways that their impairments otherwise make inaccessible. Beginning in the 1960s, researchers have developed and built a range of SSDs, including mainly visual-to-tactile systems (e.g., the Optacon of Linvill and Bliss (1966); the Tactile Visual Substitution System of Bach y Rita *et al.* (1969)) and visual-to-auditory systems (e.g., the vOICe of Meijer (1992); the Prosthesis Substituting Vision for Audition of Capelle *et al.* (1998)).

Among the many design challenges that must be met by successful SSDs, one of the most serious is the *scope problem* — that of ensuring that the novel substituting modality has whatever representational resources are required to encode the full range of targets normally covered by the substituted modality. The solution to this problem built into extant SSDs is, roughly, to work bottom up — to preserve the representation of the basic distribution of energy to which the substituted modality is sensitive, and then to hope that all other relevant perceptually represention can be derived therefrom. Unfortunately, I will argue below, this attractively simple solution to the scope problem fails. For there appear to be properties represented by our perceptual modalities whose exemplification is not fixed by the representation of the distribution of (the relevant forms of) basic energy. Hence, a SSD whose basic representational vocabulary is limited to the distribution of such basic energy will inevitably leave things out.

Here's how I'll proceed. In §1 I'll start from a simple view of perceptual representation, explain how mismatches of representational scope create a serious obstacle for the design of SSDs, and then propose a simple view about how SSDs might solve the scope problem by working bottom up from the representation of the basic energy distribution. Then, in §§2–4 I'll consider a series of perceptual phenomena that suggest that this bottom up strategy is inadequate — phenomena suggesting that the representation of the basic energy distribution in a modality is insufficient for fixing representations of all the features made available in that modality. And in §5 I'll draw lessons about what this shows about perception and the prospects for SSDs.

1 Perception, substitution, and scope

On the construal suggested above, sensory substitution amounts to replacing a first suite of representational capacities associated with one perceptual modality — if you like, an informational channel — with a second. I think it is fair to say that this conception of SSDs is reasonably prevalent amongst researchers. Thus, for example, Ward and Meijer (2010) write that, "Sensory substitution devices artificially convert information normally delivered to one sense into a representation that is compatible with an alternative sense" (492). Or, in the words of Proulx (2010), "sensory substitution devices for the blind provide the missing visual input by providing information that another sensory modality, such as the tactile or auditory system, can process" (501).

Of course, as a general matter, it is not always easy or even possible to coerce one informational channel into serving the representational/information-carrying functions

of a second. Among many other problems, disparities in carrying capacity, interface properties, or spatiotemporal resolution may render this sort of substitution difficult or impossible. But here I want to focus on an even simpler obstacle to substitution concerning mismatch of scope.

1.1 The scope problem

It is hardly earthshaking news that different modalities ordinarily represent different ranges of properties/events: in the normal run of things (i.e., without special equipment), vision but not audition represents distal shape, touch but not olfaction represents temperature, and so on.[1]

Such mismatches of representational scope create obstacles that successful SSDs must surmount. Obviously, no SSD will count as a fully adequate representational substitute for a modality, considered as a whole, unless it succeeds in representing the full suite of properties normally in that modality's representational scope. This is a highly non-trivial requirement given that modalities have such extensive representational scopes. But even more limited SSDs that aim to replace only a subset of the representational functions of a modality — say, those that attempt to represent through auditory means the normally visually accessible features of size and form, but not color — face the challenge of taking over those intended representational functions in the wide variety of situations in which the substituted modality normally does its job. Either way, the point is that the design of successful SSDs must, on pain of failing at the task for which they are intended, somehow meet the non-trivial constraint of augmenting the representational scope of the substituting modality so as to make it include some or all of the scope of the substituted modality.

1.2 Solving the scope problem from the bottom up

There is, however, a way of thinking about perceptual modalities and their operation — one that I suspect underlies the design of many extant SSDs — that suggests a possible (and principled) way forward. This strategy is built from two thoughts.

The first is that each modality is specially connected with one or more types of distal energy to which its proprietary types of transducers are sensitive.[2] Thus, for example, visual (but not auditory, tactile, vestibular, etc.) systems have photoreceptors as their basic transducers, and these transducers are selectively responsive to a particular form of environmental energy — light intensities (or, if you like, light intensities at locations). Likewise, auditory (but not visual, pain, etc.) systems have receptors that are selectively responsive to certain kinds of mechanical pressure waves (possibly at locations). And similarly for the other modalities.[3]

[1] Of course, some properties fall in the representational scope of multiple modalities: thus, for example, some (but not all) shape properties can be represented by both vision and touch. But such properties are isolated exceptions that prove the general rule: the intersection of the representational scopes of two modalities is typically a small subset of both. While we may wish to say that such cases count as a limited, if common, form of sensory substitution, we presumably want our SSDs to be much more general in their operation — we want to build SSDs that represent a much wider range of distal targets. As such, isolated cases of scope overlap won't help with the scope problem we are now considering in anything like its full generality.

[2] This idea can perhaps be seen as an expression of the Aristotelian view that individual modalities are particularly, and essentially, bound up with their own proper sensibles/objects.

[3] Heil (1983) defends the related, but stronger view that we can individuate sensory modalities by the distal energy types to which they are responsive; it is also possible to read Keeley (2002, 6) as

The second is the thought that, though modalities have wide informational scopes, all of the properties that they represent can (in principle) be fixed by their representation of the distribution of whatever types of basic energy to which they're responsive. In other words, the view is that a modality's basic energy distribution forms a computationally sufficient basis for the derivation of its (basic and non-basic) representational scope. So, for example, though vision lacks receptors specifically and selectively sensitive to exemplifications of *squarehood*, vision manages to represent whether there are local squares in the environment because its representation of the array of light intensities (the form of distal energy to which its proprietary receptors are sensitive) is sufficient, given enough computational power, to fix whether or not there is a local instance of *squarehood*. For, the thought goes, if you know where the light is, then you can (given enough computational power) figure out where the edges are, and how long they are, and from this you can figure out where the squares are. And, indeed, though the computational story may be more complicated, we might hope to extend this picture to all of the other features in the representational scope of vision. Thus, perhaps we can derive representations of distal color by some sorts of complicated comparisons between lightness intensity representations from receptors with different spectral tunings; or perhaps we can derive representations of motion by intertemporal comparisons between representations of the distribution of form (themselves derived ultimately from representations of lightness). Similarly, the thought is that preserving the representations of where the light is would provide for our device a sufficient computational basis for representing, in principle, everything vision can represent. Finally, let us suppose that this thought can be extended to other modalities, so that we would aim to account for the representational scope of each modality in terms of the distribution of energy to which its proprietary receptors are sensitive.

As I suggested above, appeal to these two thoughts offers a hope for a principled approach to the scope problem for SSDs. For even if we can't specify in advance the full representational scope of the modality for which we are attempting to build a representational substitute, it is presumably far easier to determine the form(s) of basic energy to which its receptors are sensitive. And if we can then build an alternative channel that represents those forms of energy — one that preserves in the substituting modality the information about the basic distribution of energy in the substituted modality, we will have in place a computationally sufficient basis for representing all of the properties in the full representational scope we are aiming to capture.

Importantly, on the story we're contemplating, the wide-ranging representational scope of each sensory modality is ultimately constrained by, asymmetrically dependent on, and supervenient on its representation of the distribution of the energy to which its

advocating such a view (though he wants to add further conditions). One problem for the stronger view is that it appears to individuate modalities both too coarsely and too finely. It appears to individuate modalities too coarsely because, for example, it is unobvious that there are indeed distinct distal energy types that correspond to gustatory taste and olfaction. Similarly, the stronger view predicts that thermoregulation and vision in pit vipers comprise a single modality, since both are responsive to wavelength energy (Kardong and Mackessy, 1991). On the other hand, the stronger view also appears to individuate modalities too finely: for example, it doesn't allow for a unified modality of touch, since the latter is responsive to several distinct forms of energy including temperature, pressure, and elasticity (see Fulkerson, 2013).

Crucially, the idea expressed in the main text — to the effect that modalities are specially connected with particular types of distal energy — is unlike the stronger view in not committing to any proposal about the individuation of modalities. Consequently, it is not vulnerable to the problems posed above for the stronger view.

modality-specific transducer types are responsive. Hence it seems reasonable to call the proposal we've been discussing a *bottom up* strategy for solving the scope problem.

To make clear the just what the proposed bottom up strategy looks like and why it is attractive, it may be helpful to contrast a purely hypothetical approach to the design of a vision-to-audition SSD for shape that does not conform to it. Thus, suppose we somehow build into the device some kind of primitive *squarehood* detector that makes an auditory signal in the presence of local squares, but without doing so by way of representing the distribution of light, or edges, or any other intuitively more basic properties. If we manage to do this, we will have brought the normally visually accessible (but not normally auditorily accessible) property *squarehood* into the representational scope of our system, and so will have made an advance on the scope problem. However, because our imagined system represents *squarehood* without representing the distribution of light, edges, or the like, there is no guarantee that it will have the capacity to represent other normally visually accessible properties, such as texture, depth, illumination, motion, etc. Thus, although our system will have made an advance on a single instance of the scope problem, there are all kinds of other normally visually accesible properties — which are, of course, just other instances of the same problem — about which all bets are off.

Whereas, the bottom up strategy is one that begins by representing the distribution of light intensity, and from there computing representations of (as it might be) first lightness discontinuities, then edges, then simple shapes, and so on. Crucially, because this style of computational processing works upwards from a representation of the basic energy distribution, and on the assumption that representations of that sort form a computationally sufficient basis for the representation of not only *squarehood* but everything else in the representational scope of vision, we know that our square-detecting system will, by dint of whatever allows it to represent *squarehood*, also (in principle) be capable of representing texture, depth, illumination, motion, and so on. In other words, a system implementing the bottom up strategy won't find itself stuck with a merely partial solution to the scope problem in the way that the system employing a primitive *squarehood* detector did.

1.3 A worry: Emergent features?

Near as I can tell, some version of the bottom up strategy sketched above appears to underlie the design of all extant SSDs. And this fact should come as no surprise if, as I have suggested, that strategy offers a promising, principled, and attractively simple approach to overcoming a problem that any successful SSD must solve (viz., the scope problem).

However, in what follows, I will argue that the bottom up strategy fails. For there appears to be a range of well-known perceptual phenomena suggesting that, contrary to the explicit assumptions underlying that strategy, there are what I will call *perceptually emergent features* — perceptually represented features for whose representation it turns out that a representation of the basic energy distribution is *not* a sufficient computational basis.[4]

[4]The term 'emergence' is intended to suggest a failure of fixity from below. Crucially, the failure of fixity at issue is a relation between representations of properties, rather than properties themselves. To say that F is perceptually emergent is to say that representations of F are not determined computationally by representations of other (intuitively, lower-level) properties. This claim leaves open the question of whether F is *metaphysically* emergent — viz., whether there is a

I want to emphasize that the threat about emergent features is not simply about the *de facto* limitations on the stock of available representational vocabulary (as it were, the stock of predicates) that perceptual systems happen to use. To see what that kind of threat would amount to, imagine that we build a representational system that has internal representational states corresponding to the English words for the individual colors 'red,' 'green,' 'orange,' and the rest (viz., states that are selectively responsive to instances of *redness*, *greenness*, *orangeness*, and so on), but that lacks an internal representational state corresponding to the English word 'colored' (viz., one that is selectively responsive to instances of the relatively determinable property *colored*). Now, it is a metaphysical fact — one that is independent of the particular features of our imagined representational system — that the extensions of *redness*, *greenness*, *orangeness*, and all the other individual colors collectively fix the extension of *colored*. *A fortiori*, whatever information comprises a computationally sufficient basis for the representation of the individual colors also comprises a computationally sufficient basis for the representation of *colored*. Consequently, by fixing the distribution of *redness*, *greenness*, *orangeness*, and all the other individual colors in the scene, our imagined system simultaneously fixes the distribution of *colored* in the scene. And it does this *even if it lacks a predicate that refers to the latter as such*. Intuitively, the problem for the imagined system we are now considering is a fairly shallow matter of vocabulary limitations. It is not as if there's some extra property about whose distribution our system must remain non-commital. Rather, the problem is only that our system lacks a piece of simple vocabulary for a property about whose distribution it is indeed commital.

In contrast, the challenge emergent features pose to systems implementing the bottom up approach to the scope problem is deeper. The difficulty about emergent features is that their exemplification is not fixed by the distribution of basic energy, hence that complete representation of the distribution of basic energy is not a sufficient computational basis for deriving representations of them. And this is a problem to which the mere addition of vocabulary is not the answer. For even if a system does contain predicates for such emergent features, representation of the basic energy will, by itself, leave open whether or not these predicates apply.

I hope it is clear that emergent features, if they exist, pose serious threats to the bottom up solution to the scope problem. For, if they do exist, then this would mean that the representation of the total distribution of basic energy in a modality — what the bottom up strategy takes to be computationally sufficient for capturing the entire representational scope of its intended target — is not up to the job. And this, of course, would mean that the bottom up strategy cannot do what it promises.

2 Emergent (visual) objects

As I say, I believe there are many well-documented and well-known perceptual phenomena that suggest the existence of features that are emergent in the sense just discussed. In this and the next two sections I'll present three different representative but

---

metaphysical failure of determination of F by lower-level properties — in the sense discussed by the British Emergentists (e.g., Alexander, 1920; Broad, 1925), *inter alia*. (Though I won't distinguish explicitly between perceptual and metaphysical notions of emergence in the following, I will always mean the former.)

not exhaustive clusters of such phenomena, and discuss how they support the existence of emergent features.[5]

I'll begin to argue for the existence of emergent features by considering a first cluster of phenomena that involve, in various ways, the notion of visual objecthood. Considerations about visual objecthood seem a good place to begin both because the phenomena are familiar, and because they'll suggest morals (which I'll draw out below) about where to look for other phenomena germane to our purposes.

The crucial thought behind research on visual objecthood is that the visual system does not just represent (/carry information about) an array of exemplified attributes in the local scene. Rather, it parses the scene into individual objects — roughly, bounded and connected things that trace out continuous paths in spacetime (but that needn't retain their kind affiliations to count as persisting over time),[6] and represents attributes as being exemplified by those objects. That is, the visual system represents not just the content (as it may be) *there is redness*, but *individual a is red*.

Visual objecthood is relevant here because there are a number of (much-studied) configurations in which we have reasons for thinking that the visual system is representing objecthood, but where such representations are not plausibly derived exclusively from information about the distribution of light intensity. In other words, these configurations provide evidence that visual objecthood is an emergent feature.

2.1 Apparent motion

A first such configuration involves apparent motion (e.g., Φ- or β-movement).[7] It is well-known that, for example, a pair of successively presented static images of a line or other simple figure in different locations (presented within the right spatial and temporal ranges) will be perceived as a single object moving continuously from the first presentation location to the second, rather than as two different objects. Crucially, however, the distribution of light available to vision is by itself insufficient to compel the single object interpretation that is in fact adopted by the visual system: the observed light distribution (at the sampling rate available to vision, in any case) is compatible with the presence of either (i) a single object moving through a continuous trajectory, or (ii) a pair of distinct, static, objects presented successively.

It appears that, in such cases, the representation of object identities at work in the visual system commits it to representing the world as per (i). Hence, the visual system *is* representing object identities in a way that results in its choosing between the alternatives (as it happens, it chooses erroneously: the correct description of the world is as per (ii)). The present point is that, because the light distribution caused by situation (i) is identical to the light distribution caused by (ii), its representation of object identities cannot have been fixed by its representation of the light distribution. Since object identity is both represented by the visual system and not fixed by a representation

[5]The examples below are mostly, though not wholly, taken from vision; hence, standard anti-visuocentric warnings about generalizability to other modalities apply as usual. (I should also mention that Burnston and Cohen (2013) appeal to several of the phenomena about objects discussed below for a different reason — namely, to argue against the view that feature representations are uniformly prior to object representations.)

[6]For one influential account of just what visual objects amount to, see Spelke (1990). What I say below, however, is agnostic about the details of Spelke's view.

[7]Matthen (2005, 278ff) also discusses apparent motion cases as evidence for something like what I am calling emergent features.

of the complete distribution of light (i.e., that basic form of energy to which visual receptors are sensitive), it is emergent in our sense.

2.2 Object tracking through occlusion

A second type of configuration we can use to make the same sort of point arises in experiments involving multiple object tracking. In these experiments, researchers have shown that subjects can successfully track four to five objects as they move randomly and independently among a larger set of qualitatively identical, also randomly and independently moving items.[8] Interestingly, subjects can continue to track objects when, during the motion phase, they appear to move behind and remerge smoothly from opaque occluders (and even invisible occluders) on the screen (Scholl and Pylyshyn, 1999).

Now, as we remarked about the apparent motion cases, the distribution of light during the temporal interval through which an object appears to move behind and then reemerge from an occluder is consistent with either of two possibilities. Either (i) there is a single object that is moving behind the occluder, following a continuous trajectory, and then reemerging, or (ii) a first object moves behind the occluder and stops, at which point a second, distinct object reemerges.[9] However, and also as in the case of apparent motion, we have reason to believe the representations in the visual system effectively reflect a choice between these open alternatives. For in order to carry out the tracking task successfully (as it does), the visual system must be representing object identities in a way that amounts to a choice in favor of the one-object interpretation, (i). Again, it appears that object identity is both represented by the visual system and not fixed by a representation of the total light distribution: it is emergent.

2.3 Feature binding

At a first pass, the moral I have been drawing from the cases of apparent motion and tracking through occlusion discussed above is that a representation of the light distribution underdetermines the representation of objects carried by the visual system. Consequently, if there are additional represented features that depend on visual objecthood, then we should be unsurprised to find that they, too, are underdetermined by a representation of the light distribution. I claim this is exactly what we find.

To see this, consider the visual representation of binding relations between (other) visually detected features. Binding relations are those relations we represent between two features when we visually represent them as holding of one and the same individual. This situation is unremarkable in visual perception. When we visually represent a red triangle, as it might be, we are not just representing that there is something red and that there is something triangular: we are representing that there is some one thing that exemplifies both *redness* and *triangularity*. As Clark (2000) argues persuasively, the visual system must routinely represent binding relations, or

[8] In a typical setup, the objects are presented before the motion starts, and a subset of them flash briefly to signal that those are the ones the subject should track. All the objects then move randomly and independently around the screen for a while, and then come to a stop, at which point the subject indicates which of them she had been tracking. For an overview of this research, see Scholl *et al.* (2001); Pylyshyn (2003, 2004).

[9] It turns out that qualitative identity is not necessary to sustain the single object interpretation in either apparent motion (Kolers and Grunau, 1976) or tracking through occlusion (Scholl *et al.*, 1999; Bahrami, 2003).

else we would be unable to distinguish visually (as we obviously can) a stimulus consisting of a red triangle and a blue square from a stimulus consisting of a red square and a blue triangle. To make this distinction (which amounts to solving the Many Properties Problem of Jackson (1977)), the visual system must represent binding relations between color and shape properties — it must group one color with one shape, and the other color with the other shape. And as Clark goes on to show, the most plausible explanation of visual binding rests on the idea that the visual system represents features as being exemplified by individuals. That is, it groups *redness* with *triangularity* by predicating both of one individual, a, and groups *blueness* with *squarehood* by predicating both of a second individual, b.

It is plausible, however, the individuals in terms of which binding is implemented are just the visual objects we have been discussing above.[10] But if, as we have seen, representing the light distribution underdetermines visual objecthood, and if visual objecthood is needed to fix the visual representation of binding relations between features, then representing the light distribution will also underdetermine the visual representation of binding relations. (In effect, this is just to say that representation of the light distribution is by, itself, unable to distinguish spatiotemporal featural overlap from feature coinstantiation.)

It would seem, then, that interfeatural binding relations are both represented by the visual system and not fixed by the representation of the distribution of light. And this is to say that, as before, we have uncovered an instance of emergent representation in the visual system.

2.4 Non-visual objecthood

A final consideration about the relevance of objecthood to our purposes rests in the observation that modalities differ in whether and how they represent objecthood in the first place.

As we have seen, a range of converging considerations support the view that vision parses the distal environment into functional units, or objects, which are then singled out for special treatment by the visual system — e.g., they exist simultaneously, persist in time, serve as loci for the exemplification of visually detected features, and facilitate feature binding, completion, tracking, and figure-ground segmentation in vision. Several authors have argued that there is something similar going on in audition — that the auditory system bundles bits of the environment into simultaneously existing and persisting units that serve analogous functional purposes within audition (Bregman, 1990; Kubovy and van Valkenburg, 2001; O'Callaghan, 2008; Matthen, 2010; Nudds, 2010). And perhaps, as some have claimed, some similar things can be said

[10] I say this view is plausible, but there are two or three alternatives (depending on how you count) in the running that deserve motion. On one such alternative, which Clark (2000) defends eloquently, sensory individuals are locations (/regions) in the space around the perceiver. On another, defended by Brovold and Grush (2012), sensory individuals are "gobjects," which are "anything isolated by gestalt criteria" (11).

Briefly, I doubt the location view can be a correct general account of perceptual individuals, since there exist cases in which the visual system represents multiple individuals at one location (Blaser *et al.*, 2000), and cases in which the visual system represents one individual at multiple locations (ordinary cases of amodal completion); for further discussion see Cohen (2004); Clark (2004). I am unconvinced that the gobject view is an alternative to the visual object view, as opposed to an implementation of it. (In any case, the difference between the gobject and perceptual object views won't matter to the present discussion — proponents of gobjects are invited to substitute in their preferred understanding of individuals as needed.)

in favor of the centrality of objects in touch perception (Klatzky and Lederman (1995, 1999); Pawluk *et al.* (2011); Fulkerson (2013, ch. 3)). On the other hand, it is much less clear that perception is organized in the same ways or to the same extent around objects in olfaction (but see Batty, 2010; Gottfried, 2010) or gustation.

This sort of variation between modalities raises additional obstacles for the bottom up strategy for the design of SSDs under consideration: it provides another reason why preserving a representation of the basic distribution of energy is, in general, insufficient by itself to solve the scope problem that the strategy was enlisted to solve. For if the substituted modality represents objects and the substituting modality does not, then any SSD mediating between the two — preserving representations of the distribution of basic energy or not — will (without further supplementation of its representational capacities) inevitably leave out perceptual information about objects that plays important roles in the substituted modality. Fundamentally, if the substituting modality lacks the capacity to bundle the basic energy into object representations as the substituted modality does, then preservation of representations of that basic energy will not, by itself, confer upon the substituting modality the ability to represent the world in the way that the substituted modality can.

Moreover, there may be representational mismatches even for SSDs mediating between substituted and substituting modalities that are both object-representing. For it may turn out that the cues to objecthood in the substituted modality are unrelated to — or at any rate come extensionally apart from — the cues to objecthood in the substituting modality. If so, then preserving representations of objecthood would require bridge principles to trigger object representations in the substituting modality on the basis of appropriate cues in the substituted modality. Again, this is not something that we can assume will fall out simply by ensuring that the SSD preserves a representation of the distribution of basic energy in the substituted modality.

All of this is to say that perceptual objecthood is emergent in our sense.

3 Emergent endogenous contributions

The examples of emergence discussed in §2 all concerned, in one way or another, objecthood in vision. It is not too hard to come up with a high-level diagnosis of why phenomena involving visual objecthood reveal emergence: objecthood appears to be a (partly) endogenous component of the total representational package generated by perception. It is not the world by itself, but (at least partly) the collective operation of perceptual mechanisms that is responsible for chunking the world into the the units that perception cares about and represents. In contrast, the total array of basic energy is exogenous. The constraint we have been examining — viz., the constraint of preserving the representation of the basic energy distribution — is a constraint on preserving the representation of something exogeously fixed. On reflection, it is unsurprising that a device that respects that constraint, and so represents what is exogenously fixed, can fall short of fixing the (not exogenously fixed) representation of objects.

Viewing the problem in these terms suggests that we might find other cases of emergence by looking for other cases of endogenous contributions to perceptual representation. In this section, I'll carry out this search by considering two other families of cases where we have reason to think perception makes use of such endogenous contributions: the first involves classic visual illusions, while the second involves what have come to be called "natural constraints" discussed in computational theories of various sorts.

3.1 Emergence in falsidical perception: Evidence from illusion

Consider what goes on in a standard instance of a visual illusion, such as results from the interaction of the perceptual system with the Müller-Lyer configuration. Here there is a basic distribution of light energy; and, as usual, that energy causally affects visual transducers, which in turn initiate further causal transactions within the visual system; and the end result is that, again as usual, the visual system represents a range of features of the distal stimulus.

But what makes the case interesting — what makes it a case of illusory perception — is that the visual system ends up in a state that represents (falsidically) *more* than what is fixed by the properties of the mind-independent, distal stimulus. The testimony of perception to the contrary notwithstanding, the lines in the Müller-Lyer configuration are the same in length. But this means that what the energy distribution resulting from the Müller-Lyer configuration provides as an exogenous contribution to the eventual perceptual state cannot possibly include the information that the lines are of different lengths. On the contrary, the perceptual representation of a difference in length must be endogenous in origin — it is, as it were, something added by the perceptual system as a downstream result of its interaction with the impinging energy array.[11] As such, we should not expect that a SSD that preserves under substitution the representation of the exogenously fixed distribution of basic energy will, *ipso facto*, preserve the endogenously generated representation of a difference in line length in the Müller-Lyer configuration.[12] And what I have just said about the Müller-Lyer configuration obtains, *mutatis mutandis*, for many other cases of illusion.[13] If so, then such illusions can plausibly be thought of as giving rise to further instances of perceptual emergence, hence as reasons to doubt the sufficiency of the bottom up strategy for the design of SSDs.

(Something closely analogous to what I've said about illusory perception can be said as well about perception of bi- or multi-stable configurations, which is not obviously correctly described as illusory. In perceiving a Necker cube or a more complex configuration such as that depicted in figure 1, the perceptual system (at a time) "chooses" from between one of multiple interpretations left open by the stimulus configuration. In so doing, what it represents goes beyond what is fixed by the representation of the basic energy array. For on the one hand, to say that the figure is multiply stable is just to say that the representation of the basic energy distribution, as such, is compatible with multiple available representational disambiguations. And on the other hand, in favoring one of the available disambiguations, the perceptual system is representing something that is incompatible with the others. Once again,

[11] This way of thinking about illusion echoes (albeit schematically) familiar Kantian themes.

[12] I am not claiming that the representation of a difference in line length (or, more generally, that illusory representations) cannot be preserved under sensory substitution. I am claiming, rather, that preserving a representation of the basic energy distribution alone is insufficient to guarantee this. See below for futher discussion.

[13] Though not all. The famously illusory representation of the shape of a half-immersed oar, for example, plausibly *is* fixed by the distribution of basic energy that strikes the retina. That is because this illusion is the result of a refraction of light that occurs prior to its impinging on perceptual systems at all. That said, I take it that very many classical illusions (such as the Müller-Lyer) are unlike this case in that, as claimed in the main text, they depend crucially on endogenously determined aspects of the operation of the perceptual system. In the discussion of illusion that follows I mean to be focussing only on instances that are in this respect like the Müller-Lyer case rather than the half-immersed oar case. (Thanks to Fiona Macpherson for raising this qualification.)

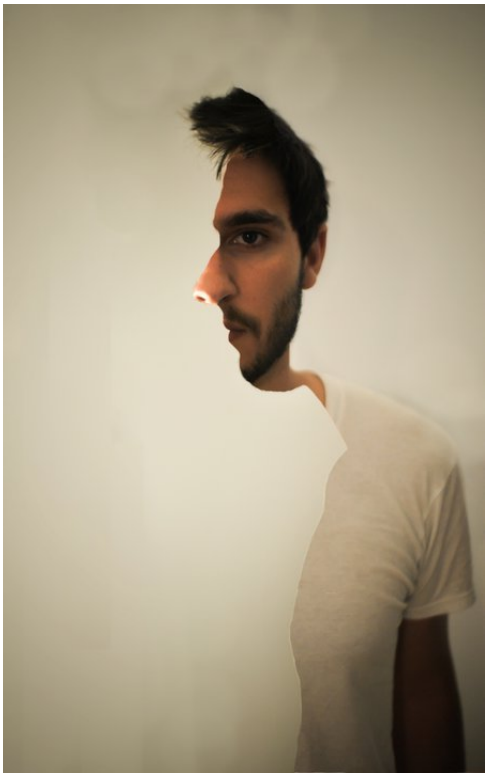

Figure 1: In representing a bistable figure in one rather than another of the ways left open by the distal configuration,what the perceptual system represents goes beyond what is fixed by the distribution of light intensity available to its receptors. Image © Mike Edmonds Photography; used by permission.

therefore, we should not expect that a SSD that preserves under substitution just the representation of the exogenously fixed basic energy distribution will preserve the endogenously generated representation that corresponds to one disambiguation of the configuration over the others.)

Now, the latter prediction is *prima facie* at odds with at least some of the claims about illusion under SSD that have appeared in the literature. Unfortunately, however, the evidence is less straightforward than it might initially appear.[14]

One of the first reports on this matter comes from the personal reflections of Gerard Guarniero (1977b), a congenitally blind Ph.D. student in philosophy who served as a subject for Paul Bach y Rita's work with the TVSS, and who went on to write a dissertation on space perception (Guarniero, 1977a). Guarniero self-reports that some classical illusions are indeed preserved under TVSS (cf. Bach y Rita, 1984). And these reports are cited approvingly by Noë and O'Regan (2002), who remark that "TVSS-aided perception is liable to familiar forms of visual distortion and illusion, e.g., distant objects 'look' small, objects can occlude each other, etc." (19).

Despite such claims, however, it turns out that Guarniero's self-reports do not obviously support the idea that visual illusions (at least in the sense considered above) are preserved under sensory substitution. What Guarniero reports is that, after training and adaptation to the TVSS, he gained fluency at estimating distal properties of objects from proximal representations — that is, he acquired skill at estimating object size from proximal size, estimating parallels from proximal convergence, and correlating distance with proximal elevation. I want to claim that, though these are potentially interesting reports that may bear on important questions about perception (e.g., about just how to understand such so-called proximal representations, and their relations to so-called distal representations; cf. Cohen (2010, 2012)), they are not directly relevant to our question about the preservation of illusory representations under sensory substitution.

The first point to make in this connection is that the kinds of cases on which Guarniero reports success do not fit the usual understanding of perceptual illusions — viz., instances in which a subject perceptually represents a feature as qualifying an object that, in fact, it fails to qualify. Indeed, I take it that one main motivation for introducing proximal representations (*qua* alternative to distal/object representations) in the first place is to *avoid* the attribution of perceptual illusion: in so doing, we can say that the proximal representation of convergence while looking at the train tracks recede into the distance is not erroneous, even though a distal representation of convergence would be an illusion in the same circumstance. But if they are not cases of illusion, then they do not speak to our question.

The second point to make is that the abilities Guarniero reports acquiring (namely, the abilities to extract distal representations from proximal representations) are plausibly understood as involving representations formed by inference rather than perception. That is, it is plausible that what Guarniero reports gaining by adaptation to the TVSS is not a capacity to form new perceptual representations, but rather a capacity to form new representations by inferring from the same set of perceptual representations initially present. Thus, relying on standard (if controversial) criteria for the perception/inference distinction, we would expect to find that the new representations Guarniero reports coming to form are cognitively penetrable and selectively impaired

[14]One problem about this evidence is that there is surprisingly little of it. Given the thousands of known illusions, one might have expected to find large numbers of published papers investigating whether and under what conditions such effects persist under SSD. That does not seem to be the case.

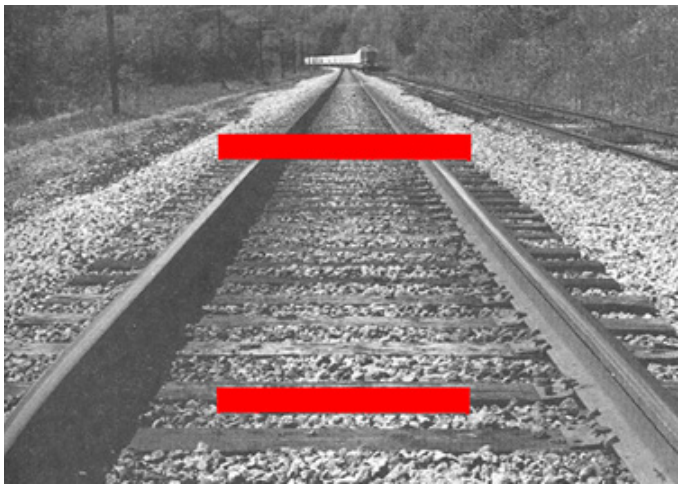

Figure 2: The Ponzo illusion is a visual illusion that occurs when parallel horizontal lines of equal length are displayed over a pair of oblique lines that converge toward the top of the display; the classic result is that subjects overestimate the relative length of horizontal lines as these are placed higher in the display.

with the capacity for general inference. In contrast, the representation resulting from a classic illusion configuration, such as the representation of the difference in line length in the Müller-Lyer configuration, is a parade case of a perceptual representation: it is cognitively impenetrable and spared by lesions that impair the capacity for general inference.

What all this suggests, then, is that Guarniero's reports are plausibly consistent with my claim that illusory representations, such as those present in classic visual illusions, need not be preserved under SSDs that are designed to preserve the representation of the distribution of basic energy in the substituted modality.

In a more recent pair of studies that address our question, Laurent Renier and his colleagues investigated susceptibility to the Ponzo illusion (depicted in figure 2) and the vertical-horizontal illusion (depicted in figure 3) in subjects using a prosthesis for substitution of vision with audition (PSVA) Renier *et al.* (2005, 2006). These investigators were able to induce both the Ponzo and vertical-horizontal illusions (in some subjects) under PSVA. But, again, the situation is more complicated than this lets on.

To begin, Renier *et al.* (2005) report that, initially, their subjects were not susceptible to the Ponzo illusion under PSVA at all, because they were able to perform the line length estimation task under PSVA without attending to representations of the converging oblique lines crucial to the visual version of the illusion. They found that, by requiring subjects using PSVA to consider the two oblique lines of the stimuli before comparing the length of the two horizontal bars, they could make the illusion reemerge. But even under this condition, the effect remained weak: they found the effect present in sighted subjects (but still smaller in magnitude than among control sighted subjects not using PSVA), but not at all in their "early blind" subjects (subjects blind before their

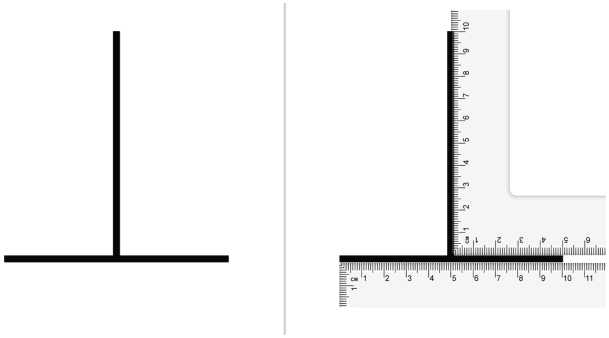

Figure 3: In the vertical-horizontal illusion, subects systematically overestimate the length of vertical lines relative to horizontal lines of equal size.

20th month of age). And Renier *et al.* (2006) report a similar pattern for the vertical-horizontal illusion: the effect was strongest in sighted control subjects not using PSVA, somewhat weaker in blindfolded sighted subjects using PSVA, and completely absent in early blind subjects using PSVA.

I want to stress three lessons about these results.

A first lesson, which emerges from the finding that the investigators failed to replicate the Ponzo illusion under substitution before modifying the task to fix attention on the oblique line elements of the stimulus configuration, is that a substitution that preserves the representation of the distribution of basic energy does not, *ipso facto*, preserve the distribution of perceptual attention. The second lesson is that, since there are aspects of the perceptual representation (here, but plausibly ubiquitously in perception) that depend crucially on which bits of the stimulus are attended, the non-preservation of the distribution of attention means that preserving the representation of the distribution of basic energy is not guaranteed to fix every aspect of the eventual perceptual representation. Specifically, it appears in the case of the Ponzo configuration that attending to the oblique lines crucially interacts with subjects' representation of the length of the horizontal lines. Hence, if a subject confronts an analogue of that configuration in another modality, but fails to attend to just the elements she would have attended to in a visual presentation, she may not go on to represent the configuration as she would have in a visual presentation. Third, and finally, the results with both the Ponzo and the vertical-horizontal configurations show that even if both the representation of the energy distribution and the distribution of attention are preserved, the illusory representation can *still* fail to show up at all (as here with the early blind subjects), or may show up in a weakened form (as here with the sighted subjects). And this is just to say that, as predicted above, preservation of the representation of the basic energy distribution under substitution does not *ipso facto* preserve the illusory feature representations: the latter are perceptually emergent in the sense described.

It seems to me, then, that (advertising to the contrary notwithstanding) the relevant extant evidence concerning classical illusions provides further support for treating the latter as involving perceptual emergence. Of course, the claim I am making is not that no illusory representations of the sort revealed in classical illusions could be preserved

under substitution. Rather, I am claiming (i) that not all such illusory representations will be preserved under substitution, (ii) that apparent preservation of such illusory representations may, on further inspection, turn out to be the result of inference, so not instances of preservation of perceptual representation after all, and (iii) that where such illusory representations are preserved, the mere preservation of the representation of the basic energy distribution is insufficient to guarantee this outcome. If these claims are correct, then cases of illusion give us further evidence that the bottom up strategy for SSDs cannot succeed by itself.

3.2 Emergence in veridical perception: Evidence from computational theory

Although perceptual illusions provide evidence of endogenous contributions to perceptual representation (and, consequently, evidence of perceptually emergent feature representations), it would be a mistake to assume that such contributions are limited to cases of falsidical perception. One way to see this is by consideration of the sorts of strategies that underlie many of the computational theories of perception that have emerged over the last several decades, which indicate that endogenous contributions play indispensable, and utterly routine, roles in veridical perception as well.

Computational theories are, of course, devoted to describing the computational processes by which perception derives and represents distal features from the initial energy distribution. Theorists in this tradition often characterize the perceptual processes they describe as solving "inverse problems" — problems of recovering rich information about the distal world from relatively sparse input. These inverse problems are challenging because, in general, the perceptual input underdetermines their solutions; hence researchers will say that the problems are "ill-posed," or provably unsolvable without imposing further constraints.

For a representative example, consider the problem of computing a visual representation of 3d form. It is an elementary fact of projective geometry that a static 2d array of incident light intensity arriving on the retina could have been produced by infinitely many distal 3d forms. Therefore, a form extraction algorithm that restricts itself to computing over input available in the retinal intensity array will be, for reasons of principle, unable to decide between infinitely many alternatives compatible with its evidence. Or, again, consider the problem of computing the surface description of surface color (say, represented as a surface spectral reflectance distribution) from the retinal light intensity array. Since the light arriving at the retina is the result of an interaction between both the illumination incident on the distal surface and the properties of the surface itself, it is a result of elementary arithmetic that the retinal evidence, taken by itself, is ambiguous between infinitely many different descriptions of the surface. Consequently, an algorithm for the derivation of surface descriptions that restricts itself to the evidence available in the retinal intensity array will, as a matter of simple arithmetic, be unable to arrive at a determinate answer. And this situation is entirely typical.

Now, of course perception *does* solve these computational problems: it *does* arrive at a determinate description of 3d form, surface color, etc. But if these problems are both solved by the perceptual system and provably unsolvable without imposing further constraints beyond those provided by the impinging energy, it follows that perception must be imposing further constraints — viz., constraints not available as such in the exogenous basic energy distribution. For exactly this reason, computational theorizing

about perception standardly proceeds by appealing to further constraints that are not exogenously fixed by the distribution of basic energy, but that instead take the form of natural, internalized regularities about the distal world.

An early, vivid, and representative illustration of this explanatory strategy arises in the classic structure from motion theorem of Shimon Ullman (1979).[15] Ullman was interested in the question of how the visual system derives representations of three-dimensional object form from a series of static two-dimensional presentations of multiple dots corresponding to locations on object surfaces. He showed that the visual system can derive the three-dimensional form of a moving object from three static two-dimensional views of four non-coplanar points.[16] Crucially, however, Ullman's derivation depends on the further assumption — not encoded in the exogenous input to vision, *a fortiori*, not encoded in the representation of the total light distribution — that the perceived object undergoes only rigid transformations in space (rotations, translations, reflections, and combinations thereof).

The assumption that moving surfaces are transformed rigidly is an instance of just the sort of "natural constraint," (/"internalized regularity") we have been discussing. That is, it is a generally (though not perfectly) reliable generalization about the perceived world, without reliance on which the input to the perceptual system would underdetermine the solution to the perceptual problem of reconstructing relevant features of the distal world, but with which the exogenous input does permit a solution. Ullman proposed that we should understand the ability of the perceptual system to solve the problems under study as resulting from its having internalized these constraints, presumably as a result of natural selective pressures imposed in an adaptive environment in which they are generally true.

Importantly, treating this and other such constraints as internalized in this sense should not be understood as claiming that they are explicitly represented in closed form in the perceptual system (Pylyshyn, 2003, ch. 3). Rather, the idea is that the perceptual system acts (i.e., carries out its characteristic computations from its input) by default in accord with these constraints in something like the way that the planets act by default in accord with Kepler's Laws.[17]

But if perception does make use of such natural constraints in the way suggested by this family of computational theories, then this gives us further reason for believing

[15]For further classic articulations of the role of natural constraints in computational accounts of perception (particularly in vision), see Marr (1982); Richards (1988); Shepard (1987, 1994, 2001). In hindsight, it is also possible (if anachronistic) to view Gestalt contraints such as "good figure" and "simplicity" (Koffka, 1935) as natural constraints in this sense.

[16]The computation assumes that the visual system has already solved the "correspondence problem" — that it has some method of associating each dot in the ith display with a particular dot in the jth display, and so treating one as an update of the very same position on the surface of the object whose form it is computing.

[17]There are, of course, theorists who posit much more by way of explicitly represented endogenous contribution in perception. Perhaps a limit case is the view of Purves and Lotto (2003), on which (roughly) perception ordinarily disambiguates the underconstrained exogenous input by appeal to an endogenously supplied and explicitly represented probability distribution over alternative hypotheses about the world. If some such view is right, then there are two distinguishable forms of emergent information in perception. First, there are the endogenous and explicit representational elements themselves. And second, there are the representations of distal features that perception derives from both the latter endogenous elements and the exogenous basic energy distribution. On the view under consideration, perception represents both of these, but neither is fixed by the representation of the basic energy distribution; therefore, both are perceptually emergent.

in perceptually represented emergent features, so for rejecting the bottom up design strategy for SSDs. For, if so, then distal features such as form, color, shape, and the like are represented by perception, although the representation of the basic energy distribution is an insufficient computational basis for their extraction. Hence, such features of the distal world are perceptually emergent in our sense.

4 Emergent high level features

In §§2–3 we reviewed two families of motivations for believing that perception represents emergent features. In this section I want to consider a third family of cases that support the same conclusion, here involving the representation of what I'll call "high-level" features. While I don't have any principled way of demarcating such features, what I have in mind are features that play a special role in perception, that are categorical, that may have particular adaptive significance, and that typically mediate other perceptual representations, e.g., by automatically attracting perceptual attention (in a way that is unlikely to be preserved under substitution). Moreover, representations of the features I have in mind lack the standard earmarks of having been derived via inference: they persist despite changes in belief, they arise effortlessly, automatically and involuntarily, and universally — viz., in perceivers from different cultures and ages (importantly including infants), and independently of their performance on standard measures of general intelligence. Consequently, we have reason for treating the representations in question as perceptual. I claim that there are several high-level features of this sort that are plausibly perceptually represented, but whose perceptual representation is not fixed by the representation of the complete distribution of basic energy available to perception — including facehood, biological motion, animacy, and perhaps causation. If I am right, then these features, too, are perceptually emergent.[18]

[18]Three concessive qualifications are in order.

The first concessive qualification is that the cases of perceptually emergent properties considered below are intended to be illustrative but not exhaustive. Since I am only trying to argue that there are some such cases, I am happy to allow that there may be many more such examples.

The second concessive qualification concerns the relation between the arguments that follow and related arguments of Susanna Siegel. In a series of important works, Siegel (2006, 2009, 2010) has argued for the related view that several high level features (her label is 'K-properties') — including *pine tree-hood* and other natural kind properties, mind-independence, and causation, *inter alia* — are represented in visual/perceptual experience. She argues for this conclusion principally by applying her "method of phenomenal contrast," which involves identifying pairs of perceptual experiences that she claims are phenomenally different from one another, and then arguing that the best explanation of this difference is that one of the perceptual experiences in the pair does and the other does not represent the target K-properties. I won't rely on Siegel's arguments in what follows, in part because I worry that even if we accept, with Siegel, that K-properties are represented in perceptual experience, we'd need further arguments to secure the claims, for which I am arguing, that the relevant high level properties are both (i) represented in the perceptual system itself (as opposed to non-perceptual elements that contribute to the representational content of perceptual experience) and (ii) not fixed by the representation of the basic energy distribution. On the other hand, if Siegel's arguments are persuasive, and if there is a successful method of extending her conclusions to the view that K-properties are perceptually emergent, then I'm happy to take the former on board as yet further reasons to believe in perceptual emergence.

The third concessive qualification is just to remark that the evidence I appeal to below is not knockdown, and certainly not deductive. Notwithstanding this remark, I take these considerations to add up, collectively, to a strong case for perceptual emergence. For better or worse, non-

4.1 Facehood

A first example of the kind of high-level feature I have in mind is the categorical feature of being the face of a conspecific.

The recognition of faces is widely believed to be treated specially by perceptual systems in a number of ways that lend credence to the idea that facehood is perceptually represented. Faces receive special visual processing not allocated to visual perception of other body parts or objects (Kanwisher, 2010; Sinha *et al.*, 2006; Sugita, 2009), and this special processing is carried out by specialized areas (Liu *et al.*, 2010; McCarthy *et al.*, 1997). Perceptual face recognition is susceptible to a characteristic overgeneration error: we (mis-) recognize a face in a cloud or a mountain far more readily than other objects. And face recognition can be selectively impaired in injuries that spare other aspects of perceptual and cognitive processing (this condition is known as prosopagnosia).

But if these considerations suggest that faces are perceptually represented, there is reason to think that the representations in question are also emergent. One way to see this point is by consideration of the so-called Thatcher effect — the finding that local anomalies in the geometric organization of faces are dramatically harder to detect in inverted than in non-inverted faces (Thompson, 2010, see figure 4).

Now, the standard explanation of our capacity to distinguish easily between the normal and anomalous images when presented right side up but not upside down rests on a representational difference. Namely, the idea is that the visual system represents the specialized high-level property of facehood in one of the right side up images (viz., the normal one) but not the other, whereas it does not represent any feature that cleanly distinguishes between the two upside down images. However, if we restrict ourselves to considering the representation of the complete distribution of basic energy available to vision, then we will be unable to distinguish between the inverted images from the non-inverted images: visual perception represents the same array of lightness intensities in the inverted and non-inverted images (modulo the global inversion that distinguishes the inverted from the non-inverted images). If this is right, then (i) there is a visual representation of the categorical feature of facehood; and (ii) that representation is not fixed by the visual system's representation of the complete basic energy distribution.

4.2 Biological motion

A second example of the kind of high-level feature in perception I have in mind, and about which many of the same things can be said, is the perceptually represented feature of being a "biological" motion. The classic finding in this area is that subjects can classify — automatically, quickly, and easily — certain but not other, equally complex, patterns of a few moving points of light as reflecting the motion of key joints in a moving organism (Johansson, 1973, see figure 5).

Once again, there are reasons for believing that this categorical feature enjoys a specialized (and possibly adaptively advantageous) role in human perceptual processing. Perception of biological motion appears to involve specialized processing (Lu, 2010), to be carried out in specialized areas (Allison *et al.*, 2000; Grossman *et al.*, 2000), and can be spared in injuries that damage gross motor and other spatial abilities (Jordan *et al.*, 2002; Kim *et al.*, 2008).

---

deductive evidence is the only kind of evidence we are going to get in such broadly empirical inquiry.

Figure 4: Local anomalies in the geometric organization of faces are dramatically harder to detect in inverted than in non-inverted faces.

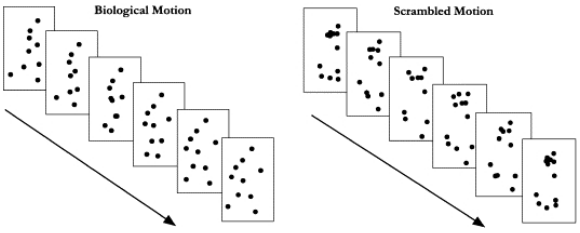

Figure 5: Subjects automatically, quickly, and easily distinguish between patterns of moving points of light that appear to reflect the motion of key joints in a moving organism (biological motion) and equally complex patterns of moving points of light that do not (scrambled motion) (Johansson, 1973).

But there are also reasons for believing that the categorical, perceptually represented feature of being a biological motion is perceptually emergent. The first arises from the mere fact that perception singles out this particular feature for quick and easy classification, while it does not appear to distinguish other, equally complex dynamic features in moving dots displays. Obviously there are differences in the representation of the total distribution of light carried by perception in biological and non-biological motion cases. But what is highly unobvious that there is any systematic difference that falls out of the representation of the total light distribution that distinguishes the two classes of cases neatly, as perception does.

A second clue to the emergence of the feature in question is that there is a Thatcher effect for biological motion as well: it is significantly easier to detect local anomalies (anomalies that perturb the visual classification of the motion as biological) in displays that are right side up than in displays that are inverted (Mirenzi and Hiris, 2011). Once again, the explanation of this difference is presumably that the visual system represents the specialized, high-level feature of being a biological motion only about the right side up and non-anomalous point light display, which is to say that there is a systematic difference between the visual representation of anomalous and non-anomalous displays when right side up, but not when inverted. But, once again, the representation of the complete distribution of basic energy available to vision doesn't provide for this kind of classificatory distinction: visual perception represents the same total array of lightness intensities in the inverted and non-inverted point light displays (even though there is a global inversion that distinguishes them). Hence, as in the case of face perception, the existence of a Thatcher effect for biological motion suggests that vision represents that high-level feature even though the latter is not fixed by its representation of the complete distribution of basic energy.

4.3 Animacy/chasing

A third example of such high-level feature in perception comes out in recent work on the perception of animacy, and particularly involving the phenomenon of perceived chasing — cases in which subjects perceptually represent a first moving shape as "chasing" a second (Gao *et al.*, 2009, 2010). This research suggests that representations of chasing behavior turn on a cluster of dynamic and configural relations between the two shapes, including the "persistence" of the chaser, overall similarity in the trajectories traced out by both shapes, and whether or not they "face" in the direction in which they move.

As Gao and Scholl (2011); Scholl and Gao (2013) argue, these representations of chasing/animacy display many of the earmarks classically associated with being perceptual. Thus, such representations are effortlessly and automatically generated, are expressed in infants as young as nine months old (Csibra *et al.*, 1999; Gergely *et al.*, 1995) and across widely differing cultures, and persist despite subjects' explicit contrary (and correct) belief that the shapes they perceive are not animate. Moreover, they are controlled by subtle visual properties of the stimulus (Scholl and Gao, 2013), can be impaired in conditions that otherwise spare motion perception (Heberlein and Adolphs, 2004; Rutherford *et al.*, 2006), and appear to implicate activity in specialized brain regions (Gao *et al.*, 2013). But, as with the other high level features we have considered, the very ease and automaticity with which the visual system recognizes chasing configurations as opposed to other, equally complex dynamic configurations represented by perception suggests that the high-level perceptual representation of

chasing is not fixed by perceptual representation of the complete distribution of light alone.

4.4 Causation

A final example of a high-level and possibly emergent categorical perceived feature is causation.[19] Famously, Michotte (4663) showed that certain configurations are spontaneously (and nearly inevitably) described by perceivers in causal terms. Thus, one class of examples Michotte discusses consists of "launching" events, in which a moving object, a, moves across a screen and comes to a halt when it makes contact with a second object, b, whereupon b begins to move in the same direction in which a had been moving. As noted, observers find it extremely tempting to describe the situation in explicitly or implicitly causal terms, perhaps saying that a pushed b or that a caused b to move.

Such representations of causation have been studied extensively in the intervening years, and the basic phenomenon — viz., that representations of causation arise in Michotte-type configurations — is now widely accepted. What is less clear is whether these representations should be regarded as perceptual. They do appear to satisfy many of the earmarks of the perceptual. Thus, Michotte himself points out that they are produced quickly, automatically, and mandatorily. Like representations of animacy, representations of causation also seem to persist across manipulations in background beliefs about causal relationships between the items (Schlottmann and Shanks, 1992). These representations of causation occur quite generally in response to Michotte-type configurations, e.g., in infants as young as six months old (Leslie, 1982, 1984; Leslie and Keeble, 1987; Oakes, 1994; Oakes and Cohen, 1990) and across diverse cultures (Morris and Peng, 1994). Moreover, they appear to be under the control of specific visual features of the displays (Scholl and Tremoulet, 2000; Scholl and Nakayama, 2002). However, such representations do not meet all of the classic marks of perceptual status that we have been drawing upon. Thus, some investigators report stable (and possibly parameterized) individual variation in expression (Schlottmann and Anderson, 1993). Moreover, it is controversial whether there is a specific neuroanatomical locus for causal perception as opposed to spatial and temporal components (Blakemore *et al.*, 2001). And, finally, so far there are no clear instances of selective impairment for this sort of causal representation (Rips, 2011).

In sum, it appears that the types of evidence we have been considering, though suggestive, fail to establish conclusively either that these representations of causation are perceptual or that they are not. However, it does seem plausible that such representations are not fixed by the perceptual representation of the light distribution alone. For the recognition of Michotte-type configurations as causal proceeds more quickly and effortlessly than recognition of other perceptually represented dynamic configurations that seem no less complex in respect of the properties of their light distributions.[20] Given these points, it seems that the right thing to say about the representation of causation is conditional: if causation is perceptually represented, then it is another instance of a perceptually emergent feature.

[19] See Danks (2009) for an excellent overview of research on these topics.

[20] This is perhaps one (anachronistic) way of expressing Hume's claim that the impressions delivered to the mind by perception fail to include anything that could amount to a source for the idea of causal efficacy.

5 Lessons

I have appealed to a range of considerations to argue for the conclusion that perception represents emergent features — that is, features whose representation is not fixed by the representation of the total distribution of basic energy. If I'm right about this, then SSDs designed so as to preserve representation of the total distribution of basic energy will not, *ipso facto*, preserve all perceptually represented features. And this means that SSDs designed in this way may nonetheless fail to solve the scope problem, and so leave out elements we may want successful SSDs to capture.

Some authors (e.g., Noë and O'Regan, 2002, 19) have had a much more optimistic view about the prospects for SSDs. They have urged that the limitations of extant prototypes are merely *pro tanto* — that they are artifacts of the crude spatial resolution and processing speed of current technology, but in principle superable with sufficient grant money and time. If the arguments of this paper are correct, however, then this optimism is misplaced, for the limitations on SSDs are more fundamental.

However, and without retracting anything of the foregoing, I want to offer several qualifications to the pessimistic assessment just offered.

First, I am in no way denying that sensory substitution is a potentially useful set of technologies. It is consistent with everything I have said that extant or future SSDs may significantly improve lives.

Second, in saying that perception represents emergent features, I do not mean to be mongering mystery about perception or perceptual processing. These representations are obviously the causal/computational outputs of various sorts of perceptual processing. The conclusion I mean to be urging is not an irrealism about all lunches, but only about lunches that are free. It is that preserving in the substituting modality the representation of the basic energy distribution in the substituted modality doesn't capture these (apparently perceptually significant) feature representations or replace the causal/computational processes leading up to them.

Third, and significantly, note that nothing in what I have said implies that SSDs *will* inevitably fail to represent what sensory modalities normally represent. It only implies that if we want them to represent what sensory modalities normally represent, then we will have to do more than preserve the representation of basic energy to which the substituted modalities are sensitive.[21] Hence, if one wants to build a system of

[21]It is perhaps worth emphasizing that this claim allows that whatever further is required (beyond preservation of the representation of the basic energy) may be something already present in our normal psychological endowment. It may be, for all that I have said, that at least some of the burden of representing perceptually emergent features is borne by amodal capacities that wouldn't need to be rebuilt for the purposes of representing those features under substitution. Thus, for example, it may turn out that the visual representation of causation is subserved partly by amodal (rather than specifically visual) mechanisms that are selectively responsive to causal relations in the world. If so, then an SSD might enlist these very same amodal mechanisms as a way of representing causation. Under the envisaged scenario, then, what it would take to ensure the representation of an ostensibly emergent feature by our SSD would not include the provision of further mechanisms above and beyond what is already present in our psychological architecture.

But this point is not an objection to the point in the main text. For if, as envisaged, the representation of emergent features under substitution is accomplished (partly) by the exploitation of standing amodal capacities that don't need to be reconstructed for use by SSDs, then one thing that such successful representation will require is preservation of those amodal capacities. Because preservation of the latter is not guaranteed by the preservation of the representation of the basic energy to which the substituted modality is sensitive, the envisaged scenario is a way of demonstrating my conclusion, rather than a way of avoiding it.

perceptual representations that includes representation of such features, then the latter will have to be built in by some other means.

Finally, I want to emphasize that it is intended as a serious, and not merely rhetorical, possibility that one might not insist on including the representation of such emergent features in one's system. For it may suffice for some purposes to allow representation of the latter to be generated by inference (or some other extraperceptual mechanism), or even to do without these representations entirely. But if one does want to include such representations within the scope of perception one is attempting to capture by means of a SSD, then the considerations adduced here suggest that mere preservation of the representation of the basic energy distribution won't suffice for that purpose. To the extent, then, that existing SSDs are constructed to serve the aim of preserving this information, their label as devices for sensory *substitution* may amount to something of an exaggeration.[22]

References

Alexander, S. (1920). *Space, Time, and Deity*. Macmillan, London.

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, **4**, 267–278.

Bach y Rita, P. (1984). The relationship between motor processes and cognition in tactile vision substitution. In W. Prinz and A. F. Sanders, editors, *Cognition and Motor Processes*, pages 150–160. Springer, Berlin.

Bach y Rita, P., Collins, C. C., Saunders, F., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, **221**, 963–964.

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, **10**(8), 949–963.

Batty, C. (2010). A representational account of olfactory experience. *Canadian Journal of Philosophy*, **40**, 511–538.

Blakemore, S.-J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A. N., Segebarth, C., and Decety, J. (2001). How the brain perceives causality: an event-related fMRI study. *NeuroReport*, **12**(174), 3741–3746.

Blaser, E., Pylyshyn, Z. W., and Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, **408**, 196–199.

Bregman, A. S. (1990). *Auditory Scene Analysis*. MIT Press, Cambridge, Mass.

Broad, C. D. (1925). *The Mind and Its Place in Nature*. Routledge & Kegan Paul, London.

[22]Many thanks to Robert Briscoe, Daniel Burnston, Alex Byrne, Kevin Connolly, Matthew Fulkerson, Cressida Gaukroger, Güven Güzeldere, Rae Langton, Don MacLeod, Fiona Macpherson, Mohan Matthen, Agustín Rayo, Laurent Renier, Barry Smith, Charles Spence, and members of the philosophy of perception reading group at the University of California, San Diego, for helpful discussions of these issues that much improved the paper. Thanks also to audiences at MIT, the Columbia-Barnard Perception Workshop, the Sensory Substitution and Augmentation Conference at the British Academy, and the SoCal Philosophy Conference, who heard earlier versions of this material and raised all sorts of useful questions.

Brovold, A. and Grush, R. (2012). Towards an (improved) interdisciplinary investigation of demonstrative reference. In A. Raftopoulos and P. Machamer, editors, *Perception, Realism, and the Problem of Reference*, pages 11–42. Cambridge University Press, Cambridge.

Burnston, D. and Cohen, J. (2013). Perception of features and perception of objects. *The Croatian Journal of Philosophy*. in press.

Capelle, C., Trullemans, C., Arno, P., and Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Transactions Biomedical Engineering*, **45**, 1279–1293.

Clark, A. (2000). *A Theory of Sentience*. Oxford University Press, New York.

Clark, A. (2004). Feature placing and proto-objects. *Philosophical Psychology*, **17**(4), 443–469.

Cohen, J. (2004). Objects, places, and perception. *Philosophical Psychology*, **17**(4), 471–495.

Cohen, J. (2010). Perception and computation. *Philosophical Issues*, **20**(1), 96–124.

Cohen, J. (2012). Computation and the ambiguity of perception. In G. Hatfield and S. Allred, editors, *Visual Experience: Sensation, Cognition and Constancy*, pages 160–176. Oxford, New York.

Csibra, G., Gergely, G., Bíró, S., Koós, O., and Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, **72**, 237–267.

Danks, D. (2009). The psychology of causal perception and reasoning. In H. Beebee, C. Hitchcock, and P. Menzies, editors, *Oxford handbook of causation*, pages 447–470. Oxford University Press, Oxford.

Fulkerson, M. (2013). *The First Sense: A Philosophical Study of Human Touch*. MIT Press, Cambridge, Massachusetts.

Gao, T. and Scholl, B. J. (2011). Chasing vs. stalking: Interrupting the perception of animacy. *Journal of Experimental Psychology: Human Perception and Performance*, **37**, 669–684.

Gao, T., Newman, G. E., and Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, **59**, 154–179.

Gao, T., McCarthy, G., and Scholl, B. J. (2010). The wolfpack effect: Perception of animacy irresistibly influences interactive behavior. *Psychological Science*, **21**, 1845–1853.

Gao, T., Scholl, B. J., and McCarthy, G. (2013). Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *Journal of Neuroscience*.

Gergely, G., Nfidasdy, Z., Csibra, G., and Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, **56**, 165–193.

Gottfried, J. A. (2010). Central mechanisms of odour object perception. *Nature Reviews Neuroscience*, **11**, 628–641.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., and Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, **12**, 711–720.

Guarniero, G. (1977a). *The senses and the perception of space*. Ph.D. thesis, New York University.

Guarniero, G. (1977b). Tactile vision: a personal view. *Visual Impairment and Blindness*, **71**(3), 125–130.

Heberlein, A. S. and Adolphs, R. (2004). Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Sciences*, **101**(19), 7487–7491.

Heil, J. (1983). *Perception and Cognition*. The University of California Press, Berkeley.

Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge University Press, New York.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, **14**(2), 201–211.

Jordan, H., Reiss, J. E., Hoffman, J. E., and Landau, B. (2002). Intact perception of biological motion in the face of profound spatial deficits: Williams syndrome. *Psychological Science*, **13**(2), 162–7.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, **107**, 11163–11170.

Kardong, K. V. and Mackessy, S. P. (1991). The strike behavior of a congenitally blind rattlesnake. *Journal of Herpetology*, **25**, 208–211.

Keeley, B. (2002). Making sense of the senses: Individuating modalities in humans and other animals. *The Journal of Philosophy*, **99**, 5–28.

Kim, J., Blake, R., Park, S., Shin, Y., Kang, D., and Kwon, J. (2008). Selective impairment in visual perception of biological motion in obsessive-compulsive disorder. *Depress Anxiety*, **25**(7), E15–25.

Klatzky, R. L. and Lederman, S. J. (1995). Identifying objects from a haptic glance. *Attention, Perception, & Psychophysics*, **57**, 1111–1123. 10.3758/BF03208368.

Klatzky, R. L. and Lederman, S. J. (1999). The haptic glance: A route to rapid object identification and manipulation. In D. Gopher and A. Koriat, editors, *Attention and performance XVII. Cognitive regulation of performance: Integration of theory and application*, pages 165–196. Earlbaum, Mahwah, NJ.

Koffka, K. (1935). *Principles of Gestalt Psychology*. Harcourt, Brace, & World, New York.

Kolers, P. and Grunau, M. V. (1976). Shape and color in apparent motion. *Vision Research*, **16**, 329–335.

Kubovy, M. and van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, **80**, 97–126.

Leslie, A. M. (1982). The perception of causality in infants. *Perception*, **11**, 173–186.

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception*, **13**, 287–305.

Leslie, A. M. and Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, **25**, 265–288.

Linvill, J. G. and Bliss, J. C. (1966). A direct translation reading aid for the blind. *Proceedings of the IEEE*, **54**(1), 40–51.

Liu, J., Harris, A., and Kanwisher, N. (2010). Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience*, **1**, 203–211.

Lu, H. (2010). Structural processing in biological motion perception. *Journal of Vision*, **10**(12), 13.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, San Francisco.

Matthen, M. (2005). *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford University Press, Oxford.

Matthen, M. (2010). On the diversity of auditory objects. *Review of Philosophy and Psychology*, **1**(1), 63–89.

Mausfeld, R. (2003). No Psychology In — No Psychology Out: Anmerkungen zu den 'Visionen' eines Faches. *Pyschologische Rundschau*, **54**(3), 185–195. English translation: http://www.psychologie.uni-kiel.de/psychophysik/mausfeld/No_Psychology_In_engl.pdf.

McCarthy, G., Puce, A., Gore, J. C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, **9**(5), 605–610.

Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions Biomedical Engineering*, **39**, 112–121.

Michotte, A. (1946/63). *The Perception of Causality*. Methuen, London.

Mirenzi, A. and Hiris, E. (2011). The Thatcher effect in biological motion. *Perception*, **40**(10), 1257–1260.

Morris, M. W. and Peng, K. (1994). Culture and Cause: American and Chinese Attributions for Social and Physical Events. *Journal of Personality and Social Psychology*, **67**, 949–971.

Noë, A. and O'Regan, J. K. (2002). On the brain-basis of visual consciousness: A sensorimotor account. In A. Noë and E. Thompson, editors, *Vision and Mind: Selected Readings in the Philosophy of Perception*, pages 567–598. MIT Press, Cambridge, Massachusetts.

Nudds, M. (2010). What are auditory objects? *Review of Philosophy and Psychology*, **1**(1), 105–122.

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, **30**, 869–879.

Oakes, L. M. and Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, **5**, 193–207.

O'Callaghan, C. (2008). Constructing a theory of sounds. *Oxford Studies in Metaphysics*.

Pawluk, D., Kitada, R., Abramowicz, A., Hamilton, C., and Lederman, S. J. (2011). Figure/ground segmentation via a haptic glance: Attributing initial finger contacts to objects or their supporting surfaces. *IEEE Transactions on Haptics*, **4**(1), 2–13.

Proulx, M. J. (2010). Synthetic synaesthesia and sensory substitution. *Consciousness and Cognition*, **19**, 501–503.

Purves, D. and Lotto, R. B. (2003). *Why We See What We Do: A Wholly Empirical Theory of Vision*. Sinauer Associates, Sunderland, Massachusetts.

Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. MIT Press, Cambridge, Massachusetts.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking (MOT): I. tracking without keeping track of object identities. *Visual Cognition*, **11**, 801–822.

Renier, L., Laloyaux, C., Collignon, O., Tranduy, D., Vanlierde, A., Bruyer, R., and Volder, A. G. D. (2005). The Ponzo illusion using auditory substitution of vision in sighted and early blind subjects. *Perception*, **34**, 857–867.

Renier, L., Bruyer, R., and DeVolder, A. G. (2006). Vertical-horizontal illusion present for sighted but not early blind humans using auditory substitution for vision. *Perception & Psychophysics*, **68**(4), 535–542.

Richards, W. A. (1988). *Natural Computation*. MIT Press, Cambridge, Massachusetts.

Rips, L. (2011). Causation from perception. *Perspectives on Psychological Science*, **6**(1), 77–97.

Rutherford, M. D., Pennington, B. F., and Rogers, S. J. (2006). The perception of animacy in young children with autism. *Journal of Autism and Developmental Disorders*, **36**, 983–992.

Schlottmann, A. and Anderson, N. H. (1993). An information integration approach to phenomenal causality. *Memory & Cognition*, **21**(6), 785–801.

Schlottmann, A. and Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *Quarterly Journal of Experimental Psychology*, **44A**, 321–342.

Scholl, B. J. and Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M. D. Rutherford and V. A. Kuhlmeier, editors, *Social Perception: Detection and interpretation of animacy, agency, and intention*. MIT Press, Cambridge, Massachusetts.

Scholl, B. J. and Nakayama, J. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science*, **13**, 493–498.

Scholl, B. J. and Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, **38**, 259–290.

Scholl, B. J. and Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, **4**, 299–309.

Scholl, B. J., Pylyshyn, Z. W., and Franconeri, S. L. (1999). When are featural and spatiotemporal properties encoded as a result of attentional allocation? *Investigative Ophthalmology & Visual Science*, **40**(4), 4195.

Scholl, B. J., Pylyshyn, Z., and Feldman, J. (2001). What is a visual object? evidence from target merging in multiple object tracking. *Cognition*, **80**, 159–177.

Shepard, R. N. (1987). Evolution of a mesh between principles of the mind and regularities of the world. In J. Dupre, editor, *The Latest on the Best: Essays on Evolution and Optimality*, pages 251–275. MIT Press, Cambridge, Massachusetts.

Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review*, **1**(1), 2–28.

Shepard, R. N. (2001). Perceptual-cognitive universals as reections of the world. *Behavioral and Brain Sciences*, **24**, 581–601.

Siegel, S. (2006). Which properties are represented in perception? In T. G. Szabo and J. Hawthorne, editors, *Perceptual Experience*, pages 481–503. Oxford University Press, Oxford.

Siegel, S. (2009). The visual experience of causation. *Philosophical Quarterly*, **59**(236), 519–540.

Siegel, S. (2010). *The Contents of Visual Experience*. Oxford University Press, New York.

Sinha, P., Balas, B., Ostrovsky, Y., and Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, **94**, 1948–1962.

Spelke, E. (1990). Principles of object perception. *Cognitive Science*, **14**, 29–56.

Sugita, Y. (2009). Innate face processing. *Current Opinion in Neurobiology*, **19**, 39–44.

Thompson, P. (2010). Margaret Thatcher: A new illusion. *Perception*, **9**(4), 483–484.

Ullman, S. (1979). *The Interpretation of Visual Motion*. MIT Press, Cambridge, Massachusetts.

Ward, J. and Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition*, **19**, 492–500.