15

Embodiment in Perception

Will We Know It When We See It?

CHAZ FIRESTONE

1 Introduction

What does it take for a cognitive process to be embodied? In a series of recent articles, Alvin Goldman has laid out a new and unifying approach to this question, centered on a refreshingly straightforward account of when cognition is and isn't embodied. On this so-called "moderate approach" to embodiment (Goldman 2012, 2013; Goldman and de Vignemont 2009), a cognitive process is embodied if and only if it recruits a body-representing system or "bodily format" in executing a cognitive task – even (or especially) when that system has been exapted or "redeployed" from its original purpose for use elsewhere in the mind. For example, if, as some experimental results suggest, understanding action verbs (such as, "kick") exploits the motoric codes usually used for one's own actions (for example, the motor commands for kicking), then language comprehension of this sort is embodied.

Embodied approaches to cognition have touched all corners of the mind, including higher-level judgmental processes such as social evaluation (Williams and Bargh 2008; IJzerman and Semin 2010), moral reasoning (Zhong and Liljenquist 2006), and theory of mind (Goldman and de Vignemont 2009; Gallese and Sinigaglia 2011). However, a primary case study across the various discussions of the moderate approach has been the possibility that embodied influences reach all the way down to *visual perception*. For example, it has been reported that wearing a heavy backpack makes hills look steeper (Bhalla and Proffitt 1999) and that grasping a baton makes reachable objects look closer (Witt *et al.* 2005). Building on the prevailing account of these results (Proffitt and Linkenauger 2013), Goldman interprets these findings as paradigm cases of his moderate notion of

Goldman and His Critics, First Edition. Edited by Brian P. McLaughlin and Hilary Kornblith.

embodiment, taking this work to suggest that bodily formats are recruited in the visual perception of space, such that spatial extents in the world are perceived relative to – or are "scaled by" – our ability to act over them. In framing the case for embodiment in perception, Goldman's account is positioned to rewrite the traditional understanding of visual perception – which has so far demanded little consideration of such body-based factors despite being arguably the most extensively studied, comprehensively modeled, and (so we think) best understood part of the mind.

This chapter will be pessimistic about such possible rewriting. After further characterizing and reviewing the evidence for moderately embodied visual perception, I will argue that such evidence does not at all support the moderate approach to embodied cognition, even when the relevant studies and accompanying theories are taken at face value. Even if body-related factors *do* influence visual perception – and indeed even if spatial perception is sometimes "body-scaled" – I will show that the prevailing theories of such body-based influences understand these effects in ways that exclude them as instances of moderately embodied cognition, because they turn out not to recruit "bodily formats" of the sort envisioned by the moderate approach. However, I'll further suggest that this result should not be surprising, and may even be embraced by proponents of this approach: by the moderate approach's own motivating principles, visual perception was a poor candidate for embodiment in the first place.

2 The Moderate Approach to Embodied Cognition

At the center of the moderate approach to embodied cognition is the notion of a "bodily format" (or "B-format" for short). B-formats are meant to be the proprietary codes used by various body-representing systems, such as the interoceptive systems that underly the proprioception of limbs and the monitoring of metabolic states, or the motor systems that generate commands for executing bodily actions. For a rough example, consider what distinguishes seeing the position of your arm from internally *feeling* (that is, proprioceptively perceiving) the position of your arm; though both states represent the position of your arm, the latter does so through a body-oriented system. Similarly, desiring to move your arm differs in at least one way from actually issuing a motor command to move your arm in that the second case involves an instruction from the motor system, and this instruction is plausibly delivered using a code specific to the motor system. On the moderate approach, any instance of cognition that recruits such body-representing systems for use in a cognitive task is an instance of embodied cognition.[i]

This account has the consequence that internally sensing the positions of our limbs and executing motor actions are themselves instances of embodied cognition (though perhaps "trivially" so). The more interesting – and more controversial – cases are those said to occur when such body-representing systems are "exapted" or "redeployed" for other cognitive tasks that might otherwise not have involved the body at all. For example, in the earlier-mentioned case where motor representations of kicking might be recruited to assist in understanding the word "kick," the same system that generates motor commands for kicking is redeployed to help out with understanding language. In that case, "action concepts may be stored in a motoric code" (Goldman 2012: 76), such that the concept tokened by the word "kick" is *that motor instruction I give my legs when I kick*.

Why should cognition work this way? The primary motivation emphasized by Goldman is that evolutionary pressures favor the reuse of pre-existing cognitive and neural resources for later-evolving capacities (see also Anderson 2010). For example, the capacity to understand action verbs evolved much later than the capacity to actually plan and execute such actions, so it stands to reason that conservative evolutionary processes may have engineered the more recently evolved function (in this case, understanding language) to exploit the resources supporting the evolutionarily earlier function (in this case, executing actions). Indeed, neuroimaging data show that hearing the word "kick," as opposed to "lick," more strongly activates the somatotopic areas of primary motor cortex associated with the legs (Pulvermüller *et al.* 2005), as if such sensorimotor areas contribute to the processing and understanding of the corresponding words. Additionally, meta-analyses of fMRI studies show that more recently evolved capacities activate more widely distributed constellations of brain areas than do evolutionarily older capacities (Anderson 2007), perhaps because the more recently evolved capacities are more likely to find, and exapt, useful structures and resources already present in the brain.

Goldman's moderate approach has already been influential in its short life (see, for example, Kriegel 2013; Shapiro 2013), and rightly so. Its core criterion for embodiment is unusually straightforward, and it does not require taking on board ancillary controversial assumptions or worldviews about the mind. It is also much friendlier to the cognitive-scientific establishment than more "radical" embodied approaches calling for the wholesale elimination of mental representations from cognitive science (for example, Chemero 2009; Wilson and Golonka 2013); true to its name, the "moderate" approach accepts that the mind is an engine of representation, and instead emphasizes the role of bodily representations in cognition. Despite this, the moderate approach remains richer than more flat-footed embodied approaches holding simply that bodily states causally influence non-bodily cognitive processing (for a review, see Wilson 2002); instead, the moderate approach holds that bodily representations often play *constitutive* roles in non-bodily cognition. Finally, in offering a unified picture of embodied cognition, the moderate approach manages to cut across other embodied approaches spanning many cognitive domains, drawing on evidence and theories from psycholinguistics, cognitive neuroscience, social cognition – and, most relevantly here, visual perception.

2.1 Embodied perception and the "scaling" hypothesis

A key contribution of the moderate approach has been to use its conceptual resources to frame the case for embodiment in *perception*, which has been a relatively neglected domain within the broader philosophical literature on embodiment. Recent experimental evidence has suggested that visual perception of the spatial environment can be altered by bodily states, such that wearing a heavy backpack makes hills look steeper (Bhalla and Proffitt 1999), holding one's arms out to one's sides makes doorway-like apertures look narrower (Stefanucci and Geuss, 2009), wielding reach-extending batons makes reachable objects look closer (Witt *et al.* 2005), and wearing specially modified shoes that make the perceiver taller makes environmental objects look shorter (Stefanucci and Geuss 2010). Individual differences in grip size and arm length also reportedly correlate with size and distance judgments of graspable and reachable objects, such that having larger limbs is associated with smaller spatial estimates (Linkenauger *et al.* 2009, 2011). Such findings sail under the flag of "embodied perception" (for reviews, see Proffitt 2006; Proffitt and Linkenauger 2013; Witt 2011).

What accounts for these results? The interpretation favored by proponents of embodied perception is that the environment is perceived relative to the body. In particular, the idea is that the body provides perception with a bounty of so-called "perceptual rulers" with which to measure up the world, such that spatial properties are represented by the visual system in units of these body-based metrics (see especially Proffitt and Linkenauger 2013; Proffitt 2013). For example, when perceiving the distance of a potentially reachable object, this approach holds that the visual system may represent the object to be at a distance equal to some multiple of the perceiver's reach; and so if the perceiver's reach varies (either between individuals, or within one individual before and after an experimental manipulation), then the perceived distance of the object in body-scaled units will vary accordingly. For instance, if the perceiver's effective reach is increased by grasping a baton, then the object will appear closer, because it will appear to the observer to be *fewer reach-lengths away*.

In framing the results this way, the perceptual-ruler approach draws on earlier foundational work on so-called body-based "scaling" in visual perception, which similarly holds that the visual system can recover spatial properties such as size and distance in units of some aspect of the perceiver's body. The most prominent such account describes cases in which visual perception represents the world relative to the perceiver's height (see, for example, Ooi *et al.* 2001; Sedgwick 1986; Wraga 1999), such that, for example, shorter observers experience objects in the world as being larger and farther away than taller observers do (for a recent philosophical treatment of these ideas, see Bennett 2011). The underlying principle, which owes its theoretical foundation to J.J. Gibson (1979), is that the "scale" of the visually perceived environment is assigned by reference to the body, and the claim from proponents of embodied perception is that height turns out to be only one of many perceptual rulers furnished by the body. This "scaling" view has been widely adopted by the broader research community working on body-based influences on perception, and it is the leading theory of such effects (see, for example, Cañal-Bruland *et al.* 2012; Glenberg *et al.* 2013; Gray *et al.* 2014; Kirsch *et al.* 2012; Lee *et al.* 2012).

2.2 Moderately embodied visual perception?

The scaling-based interpretation of the embodied perception results certainly appears congenial to the moderate approach to embodied cognition, and it is no doubt the reason that Goldman chose these results as the foundational case study for moderately embodied visual perception. Indeed, Goldman (2012) is clear that what makes these cases instances of moderately embodied cognition is not simply the experimental results themselves (viz. that grasping a baton makes objects look closer), but rather the theory built to support them (viz. that objects look closer in such circumstances because they are represented relative to the perceiver's reach). On Goldman's interpretation of the "scaling" theory, it's not just that the body causally alters perceptual processing, but rather that bodily representations are redeployed to play constitutive roles in how the visual system represents the world.

It is this core claim that I wish to challenge. Though previous work has questioned certain aspects of the evidence for embodied perception (for example Cooper *et al.* 2012; Durgin *et al.* 2009, 2010, 2011, 2012; Firestone and Scholl 2014; de Grave *et al.* 2011; Hutchison and Loomis 2006; Ontiveros *et al.* 2011; Shaffer and Flint 2011; Shaffer *et al.* 2013; Woods *et al.* 2009; for a sustained discussion, see Firestone 2013 and Proffitt 2013, as well as Firestone and Scholl, in press), we will for present purposes simply take this evidence at face value. We will assume that there are some circumstances in which modifying a perceiver's bodily states in turn modifies spatial perception – and even that this occurs by body-based perceptual "scaling." Instead, the question will be whether such evidence actually supports the moderate approach to embodied cognition. I will suggest not.

3 Visual Perception is a Poor Candidate for Embodiment

Before getting our hands dirtier with the evidence for moderately embodied visual perception, it is worth noting more generally that the moderate approach's own theoretical foundation makes visual perception an awkward choice for embodiment. The deepest motivation given for moderate embodiment in cognition has been that, as new cognitive capacities evolve over time, conservative selective pressures favor the appropriation and exaptation of existing cognitive and neural resources. Perhaps this applies satisfyingly enough to capacities such as theory of mind and spoken language (two examples favored by Goldman 2012, 2013, and Goldman and de Vignemont 2009; see also Goldman 2006), which appeared in our relatively recent evolutionary past.[ii] But this motivation also seems to predict that, all else equal, more evolutionarily ancient cognitive capacities should be less likely to admit of embodied influences. And in that case, visual perception is surely one of the poorest candidates for embodiment of any capacity in our cognitive repertoire.

Whereas spoken language may be only several hundred thousand years old (see, for example, Hauser *et al.* 2002), our evolutionary ancestors had been *seeing* for as long as they'd been doing just about anything else. Advanced eyes with image-forming lenses first developed over half of a billion years ago, and more rudimentary visual capabilities existed even before then (Land and Fernald 1992). Indeed, many paleontologists point to the evolution of such optical devices as a primary driver of the "Cambrian explosion" that transformed animal life from mostly worm-like creatures into the sundry and sophisticated forms present in modern phyla (Parker 2011). For vision in particular, then, the story is quite the other way around from how the moderate approach would have it: by its own motivating principles, the moderate approach should have placed vision near the very bottom of its list of cognitive systems likely to be embodied. And if we have been perceiving space longer than we have been grasping objects (Iwaniuk and Whishaw 2000), then it seems backwards to suggest that spatial perception "exapted" the capacity for grasping.

This larger-scale consideration is echoed by the more specific neuroscientific data that are often considered suggestive of embodiment in other cognitive domains. Recall, for example, that hearing the word "kick," as opposed to "lick," more strongly activates the somatotopic areas of primary motor cortex associated with the legs (Pulvermüller *et al.* 2005). Though the proper interpretation of such findings is quite contentious on its own (see Mahon and Caramazza 2008), it is noteworthy that analogous cases are not observed for spatial perception: there are no analogous streams of research reporting that merely perceiving the spatial properties of one's environment robustly and selectively activates brain areas for representing the body. And this is not for lack of looking: the vast literature on what is often called "scene perception" has investigated the neural bases of the perception of spatial layout, comparing experimental conditions in which subjects view images with rich 3D spatial content (such as navigable landscapes, fields, or indoor rooms) to baseline conditions in which subjects view images without such rich spatial content (such as faces, 2D collages of objects, or no image at all). However, such investigations have not implicated body-related brain areas in perceiving such spatially rich visual environments, and certainly not with the specificity apparent in studies of embodiment in language processing (for the seminal work, see Epstein and Kanwisher 1998; for a recent review, see Oliva 2013).[iii] Moreover, the diffuse patterns of activation associated with the "massive redeployment hypothesis" (for example, Anderson 2007, 2010) notably (and often explicitly) exclude visual perception: Though there is much discussion of how neural resources devoted to perception are themselves reused and redeployed in the brain, the meta-analytic data marshaled by the moderate approach to embodied cognition rarely (if ever) point to the reuse of other brain structures *for* perception.

Of course, none of these considerations *entails* that spatial perception is not embodied in the way imagined by the moderate approach. But it is well worth emphasizing that the motivations given for embodiment in general almost always fail to apply to visual perception in particular, and the moderate approach is no exception.

4 Body-Based Scaling in Visual Perception

Let us return to the empirical evidence for embodiment in perception. A wealth of research has reported that bodily states such as the size of one's hands or arms influence perception of spatial properties such as the size and distance of objects in the world, and the prevailing explanation for such effects is that the visual system "scales" the environment by bodily representations (Proffitt and Linkenauger 2013; see also discussion in Goldman 2012). In the next two sections, I'll argue that even if we assume that such bodily states affect spatial perception, and even if we further assume that they do so by body-based "scaling," we should still not consider this to be evidence in favor of the moderate approach to embodied cognition, because body-based scaling does not recruit "B-formats" in the way envisioned by the moderate approach. And in that case, no instance of such body-based scaling will turn out to be an instance of moderately embodied cognition.

4.1 The case of eye-height scaling

To make the case that body-based perceptual scaling is not moderately embodied cognition, a somewhat involved analogy will be illustrative. Whereas the particular research on embodied influences on perception has been controversial (for a review, see Firestone 2013), it is noteworthy that the "scaling" theory put forward as an account of embodied perception – which is the basis for claiming that the "embodied perception" findings are instances of moderately embodied cognition – is grounded in much less contentious work: so-called "eye-height scaling" in size and distance perception (Ooi *et al.* 2001; Sedgwick 1986; Wraga 1999).

Eye-height scaling is a hypothesized process by which the visual system is said to represent spatial properties such as size and distance relative to the altitude of the observer's eyes. (A nearby tree, say, might look 5 eye-heights high and 10 eye-heights away.) This account enjoys fairly wide acceptance (or at least little opposition) within vision science, in part because the theoretical and empirical foundation of such scaling accounts has been carefully worked out and scrutinized in the decades since they were first developed. Eye-height scaling also has more than the usual bona fides one might look for in an account of embodied cognition, having first been proposed by J. J.Gibson himself (see Gibson 1979, especially ch. 9, and the work of his student, Sedgwick 1986). This suggests a convenient way to evaluate the prospects for moderately embodied visual perception: by first determining whether the more established cases of spatial-perceptual scaling "count" as instances of embodied cognition according to the moderate approach. I will show that they plainly do not – and that it is only a short step from this result to a similar conclusion about the more recent evidence for embodied perception, which is widely assumed to operate on similar principles (Proffitt and Linkenauger 2013; Proffitt 2013).

4.1.1 *Eye-height scaling: the nuts and bolts*

We have been talking about perceiving the world "and representing spatial properties in"; what exactly does this mean, and how does the visual system achieve such body-relative visual representation? Eye-height scaling is one of many computational tricks the visual system uses to resolve ambiguity in visual information – for example, to determine from the ambiguous retinal size of an object whether it is large and far away or small and nearby. The insight of eye-height scaling accounts is to notice that, in addition to the ambiguous information in the optic array, the optical information reaching the perceiver also contains a source of *un*ambiguous size information, in the form of the visual horizon (that is, where earth appears to meet sky). The horizon appears to each observer at eye-level, which means that every observer receives information about her own height relative to the rest of the world just by looking straight ahead. Crucially, this information is invariant with the size or distance of any objects being viewed, because the horizon is essentially 'projected' onto objects in the environment: under normal viewing conditions, any object an observer sees is cut at eye-level by the horizon.

This allows the visual system to resolve certain ambiguities in visual information. Consider the observer viewing the object in Figure 15.1a. The visual system's job is to

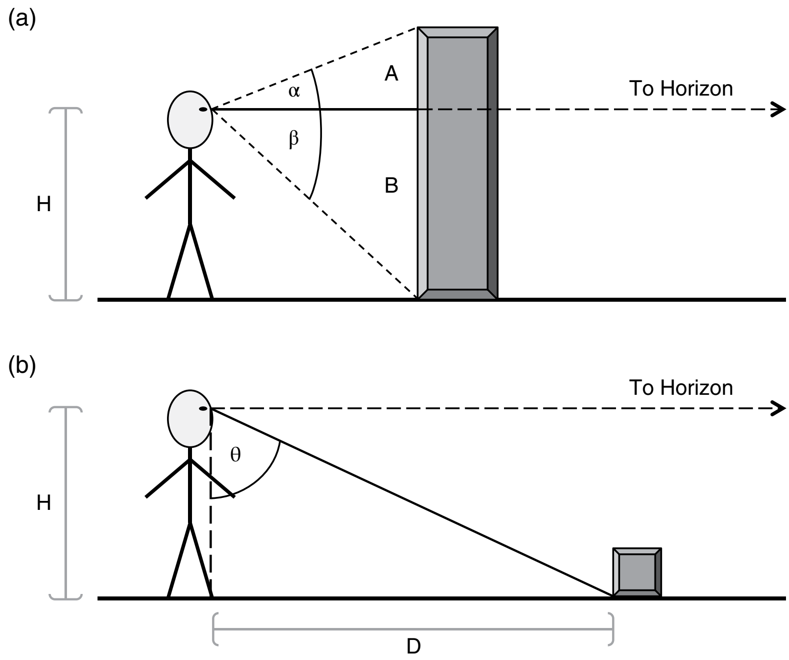

Figure 15.1 Eye-height scaling in size and distance perception.

determine the object's physical size, but the object's retinal size is ambiguous on its own; the object could be a big thing far away or a small thing close up. However, the horizon divides the object into two portions – A, the portion of the object above the horizon, and B, the portion of the object below the horizon – and the angular subtenses of these portions of the object (α and β) are related to each other in ways that track the size of the object and thus resolve the ambiguity. For example, a big thing far away is distinguishable from a small thing close up by the fact that the ratio $(\alpha + \beta)/\beta$ is greater for larger objects than for smaller objects (because α increases with the object's size, but β does not).

What does this have to do with perceiving space relative to one's eye-height? Since these relations holding between angular subtenses also hold between the sizes of the objects projecting those angles, we can set up the following equivalence:

$$\frac{\alpha+\beta}{\beta}=\frac{A+B}{B} \tag{1}$$

And since the horizon is what divides the object into A and B, and the horizon appears at eye-level, we can notice that B is always equal to the observer's own height (H):

$$\frac{\alpha+\beta}{\beta}=\frac{A+B}{H} \tag{2}$$

Or, simply:

$$\frac{\alpha+\beta}{\beta}=\frac{Size}{H} \tag{3}$$

Thus, an object's size is given simply by $(\alpha+\beta)/\beta$ (an optic-array structure commonly known as a "horizon ratio"), and the "units" associated with whatever this quantity works out to are *units of the observer's eye-height*. In other words, determining an object's observer-relative size is baked right into the process of using the horizon to disambiguate retinal size information; representing an object's size by its horizon-ratio just is computing its observer-relative size.

Eye-height scaling in distance perception is even simpler. Consider the observer in Figure 15.1b, viewing an object a certain distance away. Here, the visual system's job is to determine the object's distance, but the angular information alone is again ambiguous. As before, however, the horizon resolves this ambiguity – this time not by projecting the observer's eye-height onto the object, but instead by anchoring the angle of declination from the horizon down to the object's base. Assuming that the object is resting on the ground plane, the observer's height and the distance to the object form two sides of a right triangle whose two acute angles are the angles of inclination and declination. These quantities are all related to each other by basic laws of trigonometry, such that the distance to the object can be expressed as follows:

$$\frac{D}{H}=\tan(\theta) \tag{4}$$

In other words, distance is given simply by $\tan(\theta)$, and the units here are eye-height units. Again, the visual system disambiguates retinal information about space by representing space in units of the observer's height.

These examples should make clearer just what it means for the visual system to "scale" spatial properties relative to an aspect of the observer's body – in this case, height. And it may also now be clearer why a shorter person might be said to experience objects in the world as being "farther" away than a taller person would, or why we ourselves often observe that childhood haunts seem smaller when revisited later in life: if the visual system sometimes represents spatial properties relative to the observer's eye-height – if we can perceive sizes and distances in "units" of *copies-of-me* – then the shorter observer experiences objects in the world as being a greater number of eye-heights away or tall. This, in a nutshell, is the eye-height scaling account of size and distance perception (for discussion of this account's philosophical implications, see Bennett 2011).

Helpfully, this understanding of the general nature of body-based scaling accounts will also buy us "for free" an understanding of the body-based scaling account favored by proponents of embodied perception and moderately embodied cognition. The essential claim from embodied perception theorists (for example, Proffitt and Linkenauger 2013, who also work out the above derivation) is that, in a deep sense, embodied spatial perception *works* *like that*. And as we will now see, this understanding has the consequence that none of these cases of body-based scaling – whether by one's height or by (for example) one's arm-length – are instances of moderately embodied cognition.

4.2 Body-based scaling ≠ embodiment

Goldman (2013) has summarized his moderate notion of embodied as "the thesis that a significant amount of human cognition has its origins in representations of one's own body" (104). At first glance, the upshot of eye-height scaling looks awfully like an instance of embodied cognition on the moderate approach – so much so that one might wonder whether the moderate approach could simply declare victory on the basis of the above examples alone. Eye-height scaling is an example of how the environment is perceived relative to the observer's body, such that observers with different bodies literally perceive the world differently. Such scaling is also specific to one's *own* body, which has been an important criterion for moderate embodiment; one person's visual system cannot compute an object's horizon-ratio for some other person. Most importantly, the role played by the observer's height in influencing spatial perception in the above examples is not at all accidental, nor is it even merely causal; instead, on the sort of body-based scaling account reviewed above, representations of the observer's body play a genuinely constitutive role in spatial perception, because the cognitive processes underlying the visual system's computation of spatial properties express those properties in eye-level units. In the language of the moderate approach, eye-height scaling is unambiguously an example of how "visual representations of object size are scaled by reference to one's own bodily parts" (Goldman 2012: 82), a description Goldman uses for the flagship cases of moderately embodied visual perception.

However, there is another sense in which eye-height scaling has *nothing to do* with the body. Now that we have seen the details of how the visual system actually derives eye-height-scaled spatial representations, it is clear that the information processing that underlies body-based scaling of this sort is entirely *visual* in nature, involving only geometric transformations on optical information. Even though the end result is body-scaled size and distance, the visual system achieves this result by exploiting visual information *created by* the body (for example, the pitch of the declination angle, which is fixed by the observer's height) – not by "redeploying" interoceptive bodily representations (for example, internally sensing one's height). In fact, this sort of eye-level scaling could occur even in systems without bodily representations at all, for example if such a process were implemented in a computer vision system analyzing photographs taken from above the ground plane (see, for instance, Herdtweck and Wallraven 2013). Body-based scaling of this sort is thus "owned" entirely by the visual system: perception can and does represent space relative to the perceiver's body without employing "B-formats" of the sort envisioned by the moderate approach.

Indeed, there would be some awkward consequences for the moderate approach if body-based scaling accounts such as these were taken as instances of embodied cognition. Most acutely, large swaths of visual processing that have long sat comfortably within orthodox approaches to visual perception would suddenly (and spuriously) "count" as moderately embodied cognition as well. For example, one of the most robust visual cues to depth is the difference in an object's apparent position between the two eyes (the "binocular disparity"), which is exploited for depth perception by the process of stereopsis. (Roughly, the greater the disparity, the closer the object.) Interestingly, due to the geometry of the binocular disparity, a popular characterization of depth perception by stereopsis holds that its result is a body-scaled representation of space – in particular, a representation of space scaled by the distance between the eyes (see, for example, Howard and Rogers 1996; see also Coats *et al.* 2014.) But if even stereopsis, which has been known for centuries (Wheatstone 1838) and is as mainstream and foundational a discovery as there could be in perceptual psychology, counts as an embodied cognitive process (since it involves body-scaled representations), then the moderately embodied approach will have lost whatever value it sought to add to the relevant literatures – which was, after all, to "urge a reorientation" (Goldman 2012) in cognitive science.

At least some instances of body-based scaling, then, turn out not to require an "embodied" interpretation according to moderate approach, even though they are instances of how the visual system represents non-bodily objects and properties by reference to the perceiver's body. This is a result worth emphasizing on its own: it is a deep and subtle insight reached by body-scaling accounts that the visual system exploits information *created by* the body to represent the environment in body-relative terms, without redeploying interoceptive representations of the body. But we can also put this insight to further use, finally revealing that the theory developed to explain the "embodied perception" effects reviewed earlier (for example, backpacks making hills look steeper, batons making distances look closer, etc.) has the consequence that these phenomena too are driven by processes that are not moderately embodied.

5 When Embodied Perception is not Embodied Cognition

Equipped with the insight that the mind can and does derive body-scaled representations of non-bodily spatial extents without doing anything that should reasonably count as embodied cognition on the moderate approach, we are in a position to see how the experimental evidence and theoretical accounts presented by embodied approaches to perception also fail to be instances of moderately embodied cognition. The prevailing theory of how the sizes of our hands, lengths of our arms, and extents of our jumps (etc.) affect perception of spatial quantities is that, just as with eye-height, the visual system can scale space by those body parts. In the words of embodied perception theorists, "eye-height is not the only bodily metric used to scale space, but rather is part of a larger ensemble of perceptual rulers" (Linkenauger *et al.* 2011: 1434). Importantly, these additional body-based perceptual rulers are assumed to achieve such body-based scaling in a way relevantly similar to how eye-height scaling is achieved – by transformations that exploit regularities and patterns in the optic array, *as opposed to* the combination or integration of visual information with interoceptive body representations of the sort the moderate approach identifies with "B-formats." On this theory, the role played by our hands, arms, and legs in influencing spatial perception is essentially the same as the role height plays in influencing spatial perception in the above examples (which, we have just established, are not instances of moderately embodied cognition).

According to Goldman, what makes the "embodied perception" findings instances of moderately embodied cognition is not simply the experimental results themselves, but rather the body-based scaling theory underlying them. But in their most comprehensive statement of their views, Proffitt and Linkenauger (2013) make clear – rightly, in my view – that their scaling theory is very *opposed* to the interpretation that would be required by the moderate approach to embodied cognition. For example, they note that an earlier review (Proffitt 2006) included passages that "could be understood as suggesting that visual and non-visual information are combined in perception, thereby making perception a hybrid, consisting of information of mixed perceptual and non-perceptual origins. We do not ascribe to this hybrid view" (Proffitt and Linkenauger 2013: 171). Instead, they argue that "the visually perceived environment is fully specified by visual information," and that "visual information is not combined with, but rather is scaled by, non-visual metrics derived from the body" (171) before going through the derivation of eye-height scaling as the paradigmatic example of what they mean by this.[iv] Even by the lights of embodied perception theorists, then, these instances of body-based perceptual scaling are not instances of embodied cognition in the sense intended by the moderate approach.[v]

Moreover, it easy to see *why* embodied perception theorists interpret their data this way, rather than along the lines of the moderate approach. Looking through the collection of embodied perception studies, it is noteworthy that just about every study that manipulates the subject's body or action-capabilities also manipulates the visual information reaching the subject. In the clearest cases, this is simply because subjects can directly see the manipulated body part that would serve as the effector in the relevant action (for example, looking at one's own hand while grasping a baton before reaching for a target; Witt *et al.* 2005). In other cases, the manipulation would likely have perturbed the optic flow reaching the subject in some other way (e.g., wearing ankle weights while jumping over a gap; Lessard *et al.* 2009; see also the discussion of optic flow in Proffitt and Linkenauger 2013 and Proffitt 2013).

Perhaps most crucially, several embodied perception studies succeed despite manipulating *only* the visual information reaching the subject, without actually altering the subject's true action-capabilities or interoceptive bodily representations. For example, subjects who placed their hands in a magnifying box such that their hands looked larger (but, of course, were not in fact larger, and did not feel larger, etc.) subsequently judged graspable blocks to be smaller (Linkenauger *et al.* 2011); similarly, immersing subjects in a virtual environment and altering the depicted hand-size of their virtual avatars reportedly influenced the perceived sizes of virtual objects reached with the enlarged hand (Linkenauger *et al.* 2013).

Such results should be inexplicable on the view that the embodied perception findings are explained by the redeployment of interoceptive body representations (involving B-formats), since the interoceptive information reaching the subject was held constant across the experimental conditions, and there was thus no change in the B-formatted representations that supposedly serve as scaling metrics. Instead, if, as the moderate approach insists, we are to understand these results as involving spatial-perceptual scaling of some kind (for an alternative view, see Firestone 2013), then the only non-deflationary interpretation available is that body-based scaling of this sort operates by transformations on visual information, just as does every other sort of body-based scaling.

Thus, having previously seen that body-based scaling in visual perception does not generally require (or involve) the redeployment of B-formats, we have now also seen that the work on embodied perception – which constitutes the only evidence so far marshaled in favor of moderately embodied visual perception – shares this property, and thus that these results too should not be taken as examples of moderately embodied cognition.

6 Conclusion: Moving Forward, Moderately

I have argued that a large class of empirical and theoretical evidence that may initially appear to support the moderate notion of embodied visual perception turns out not to show (or even suggest) that vision is embodied. Is this because of some defect in the moderate approach's criteria for embodiment? I believe the answer is no, and that the moderate approach to embodied cognition remains a refreshing and potentially game-changing approach to understanding the nature and extent of body-based influences in the mind.

What went wrong, then, with the moderate approach's claims about visual perception? The core criterion for the moderate notion of embodiment – involving the recruitment and redeployment of body-representing systems for non-bodily tasks – seems right on target. Instead, I take the foregoing discussion to have shown that this core criterion is very slightly out of step with the broader spirit of the moderate approach – a spirit illustrated by what Goldman (2013) calls a "slogan" for his view: "In the beginning, what we represent is our own body" (104). As the various cases of body-based scaling in perception show, the mind constructs body-oriented representations that meet many of the in-spirit conditions for moderate embodiment put forward by Goldman, and yet do not involve the recruitment of internal, body-representing systems, and so are not instances of embodied cognition on the moderate approach.

Something has to give: either the moderate approach can abandon the claim that visual perception is embodied, or the criteria for embodiment could be tweaked so that body-based perceptual scaling *does* count as moderately embodied visual perception. It seems to me that the latter option is unadvisable: as we have seen from the example of stereopsis, inviting perceptual scaling into the moderate approach's notion of embodied cognition would have unacceptable consequences for an approach that is supposed to mark a "distinctive departure from orthodox cognitive science" (Goldman 2013: 104).

The most favorable solution, then, is to accept that body-based scaling in visual perception – of which the "embodied perception" findings are intended as an example – does not meet the moderate approach's criteria for embodiment. It is no indictment of the moderate approach that the clarity and incisiveness of these criteria helped us see that such cases are not truly instances of embodied cognition; indeed, it is clearly a strength, even if the consequence is that less of cognition is embodied than originally hoped. But this is, after all, not a wholly unexpected outcome: visual perception was an awkward case study of embodiment all along, given the moderate approach's evolutionary motivations. If the moderate approach can let go of embodied visual perception (at least on the basis of currently available evidence), then it can – and surely will – continue to make gains in the rest of the mind.

Notes

i Goldman (2013) clarifies that if "code" or "format" in this context sounds overly language-like, then "using a body-oriented format" could be understood as roughly equivalent to "recruiting a body-representing system." I will understand the notion of a "B-format" in this sense.

ii Of course, many exceptions and puzzles are lurking even in these cases. For example, if our ability to understand action verbs such as "kick" and "lick" relies in some deep way on the redeployment of neural circuitry for programming and executing such movements, then what of our apparently equal ability to understand the verbs "fly" or "slither "? Are these otherwise-similar classes of words processed by entirely different cognitive mechanisms? See Mahon and Caramazza (2008) for a discussion of these and other worries.

iii There is evidence that (e.g.) viewing images of graspable tools activates motor regions associated with grasping (for a review, see Culham and Valyear 2006). However, these sorts of results do not bear on the claim that *spatial perception itself* is embodied; instead, such results may be more suggestive that recognizing tools or planning to use them can involve simulations of their use (*modulo* the persistent obstacles in interpreting such fMRI data along these lines). Just as the claim in the "kick"/"lick" case is that merely understanding such words is embodied, the claim here is that merely perceiving the spatial properties of the environment is embodied.

iv Goldman (2012) may have even unintentionally flipped this passage's meaning in discussing the moderate approach. In quoting it, he adds the word "merely" in parentheses, such that the end result is "visual information is not [merely] combined with, but rather is scaled by, non-visual metrics derived from the body" (Goldman, 2012: 84). But I take it that Proffitt and Linkenauger simply meant what they wrote – that visual information is *not* combined with non-visual metrics derived from the body, full-stop.

v It is further telling that "motor simulation," a notion briefly flirted with by a single embodied perception study (Witt and Proffitt 2008), makes no appearance in the theoretical treatments of embodied perception (e.g., Proffitt and Linkenauger 2013; Proffitt, 2013) – especially as pertains to the "scaling" account. This continues to suggest that the redeployment of body-representing systems is not a feature of such accounts in the way that would be required by the moderate notion of embodiment.

References

Anderson, M.L. (2007) Evolution of cognitive function via redeployment of brain areas. *Neuroscientist* **13** (1), pp.13–21.

Anderson, M.L. (2010) Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* **33** (4), pp.245–66.

Bennett, D.J. (2011) How the world is measured up in size experience. *Philosophy and Phenomenological Research* **83** (2), pp.345–65.

Bhalla, M. & Proffitt, D.R. (1999) Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance* **25** (4), pp.1076–96.

Cañal-Bruland, R., Pijpers, J.R.R., and Oudejans, R.R.D. (2012) Close, and a cigar! – Why size perception relates to performance. *Perception* **41** (3), pp.354–6.

Chemero, A. (2009) *Radical Embodied Cognitive Science*. MIT Press, Cambridge, MA.

Coats, R.O., Pan, J.S., and Bingham, G.P. (2014) Perturbation of perceptual units reveals dominance hierarchy in cross calibration. *Journal of Experimental Psychology: Human Perception and Performance* **40** (1), pp.328–41.

Cooper, A.D., Sterling, C.P., Bacon, M.P., and Bridgeman, B. (2012) Does action affect perception or memory? *Vision Research* **62** (1), pp.235–40.

Culham, J.C. and Valyear, K.F. (2006) Human parietal cortex in action. *Current Opinion in Neurobiology* **16** (2), pp.205–12.

de Grave, D.D.J., Brenner, E., and Smeets, J.B.J. (2011) Using a stick does not necessarily alter judged distances or reachability. *PLoS One* **6** (2), e16697.

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., and Waymouth, S. (2009) Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review* **16** (5), pp.964–9.

Durgin, F. H., Hajnal, A., Li, Z., Tonge, N. and Stigliani, A. (2010) Palm boards are not action measures: An alternative to the two-systems theory of geographical slant perception. *Acta Psychologica* **134** (2), pp.182–97.

Durgin, F.H., Hajnal, A., Li, Z., Tonge, N., and Stigliani, A. (2011) An imputed dissociation might be an artifact: Further evidence for the generalizability of the observations of Durgin *et al.* (2010). *Acta Psychologica* **138** (2),pp.281–4.

Durgin, F.H., Klein, B., Spiegel, A., Strawser, C.J., and Williams, M. (2012) The social psychology of perception experiments: Hills, backpacks, glucose and the problem of generalizability. *Journal of Experimental Psychology: Human Perception and Performance* **38** (6), pp.1582–95.

Epstein, R. and Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature* **392** (6676), pp.598–601.

Firestone, C. (2013) How "paternalistic" is spatial perception? Why wearing a heavy backpack doesn't – and couldn't – make hills look steeper. *Perspectives on Psychological Science* **8** (4), pp.455–473.

Firestone, C. and Scholl, B.J. (in press). Cognition does not affect perception: Evaluating the influence for "top-down" effects. *Behavioral and Brain Sciences.*

Firestone, C. and Scholl, B.J. (2014) "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science* **25** (1), pp.38–46.

Gallese, V. and Sinigaglia, C. (2011) What is so special about embodied simulation? *Trends in Cognitive Sciences* **15** (11), pp.512–19.

Gibson, J. (1979) *The Ecological Approach to Perception and Action*. Houghton Mifflin, Boston, MA.

Glenberg, A.M., Witt, J.K., and Metcalfe, J. (2013) From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science* **8** (5), pp.573–85.

Goldman, A. (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, Oxford.

Goldman, A. (2012) A moderate approach to embodied cognitive science. *Review of Philosophy and Psychology* **3** (1), pp.71–88.

Goldman, A. (2013) The bodily formats approach to embodied cognition. In: Kriegel, U. (ed.) *Current Controversies In Philosophy of Mind*. Routledge, New York, NY. pp.91–108.

Goldman, A. and de Vignemont, F. (2009) Is social cognition embodied? *Trends in Cognitive Sciences* **13** (4), pp.154–9.

Gray, R., Navia, J.A., and Allsop, J. (2014) Action-specific effects in aviation: What determines judged runway size? *Perception* **43** (2), pp.145–154.

Hauser, M.D., Chomsky, N., and Fitch, W.T. (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* **298** (5598), pp.1569–79.

Herdtweck, C. and Wallraven, C. (2013) Estimation of the horizon in photographed outdoor scenes by human and machine. *PLoS One* **8** (12), e81462.

Howard, I.P. and Rogers, B.J. (1996) *Binocular Vision and Stereopsis*. Oxford University Press, New York, NY.

Hutchison, J.J. and Loomis, J.M. (2006) Does energy expenditure affect the perception of egocentric distance? A failure to replicate experiment 1 of Proffitt, Stefanucci, Banton, and Epstein (2003). *The Spanish Journal of Psychology* **9** (2), pp.332–9.

IJzerman, H. and Semin, G.R. (2010) Temperature perceptions as a ground for social proximity. *Journal of Experimental Social Psychology* **46** (6), pp.867–73.

Iwaniuk, A.N. and Whishaw, I.Q. (2000) On the origin of skilled forelimb movements. *Trends in Neurosciences* **23** (8), pp.372–6.

Kirsch, W., Herbort, O., Butz, M.V., and Kunde, W. (2012) Influence of motor planning on distance perception within the peripersonal space. *PLoS One* **7** (4), e34880.

Kriegel, U. (2013) The philosophy of mind: Current and perennial controversies. In: Kriegel, U. (ed.) *Current Controversies in Philosophy of Mind*. Routledge, New York, NY. pp.1–14

Land, M.F. and Fernald, R.D. (1992) The evolution of eyes. *Annual Review of Neuroscience* **15** (1), pp.1–29.

Lee, Y., Lee, S., Carello, C., and Turvey, M.T. (2012) An archer's perceived form scales the "hitableness" of archery targets. *Journal of Experimental Psychology: Human Perception and Performance* **38** (5), pp.1125–31.

Lessard, D.A., Linkenauger, S.A., and Proffitt, D.R. (2009) Look before you leap: jumping ability affects distance perception. *Perception* **38** (12), pp.1863–6.

Linkenauger, S.A., Leyrer, M., Bülthoff, H.H., and Mohler, B.J. (2013) Welcome to Wonderland: The influence of the size and shape of a virtual hand on the perceived size and shape of virtual objects. *PLoS One* **8** (7), e68594.

Linkenauger, S.A., Witt, J.K., and Proffitt, D.R. (2011) Taking a hands-on approach: apparent grasping ability scales the perception of object size. *Journal of Experimental Psychology: Human Perception and Performance*, **37** (5), pp.1432–41.

Linkenauger, S.A., Witt, J.K., Stefanucci, J.K., Bakdash, J.Z., and Proffitt, D.R. (2009) The effects of handedness and reachability on perceived distance. *Journal of Experimental Psychology: Human Perception and Performance* **35** (6), pp.1649–60.

Mahon, B.Z. and Caramazza, A. (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris* **102** (1), pp.59–70.

Oliva, A. (2013) Scene perception. In: J. S. Werner and L. M. Chalupa, (eds) *The New Visual Neurosciences*. MIT Press, Cambridge, MA, pp.725–32.

Ontiveros, Z., Mejia, N., Liebenson, P., Lagos, A., and Durgin, F. (2011) Cognitive feedback may cause "Tool Effects": an attempted replication of Witt (in press). *Journal of Vision* **11** (11), pp.971–71.

Ooi, T.L., Wu, B., and He, Z.J. (2001) Distance determined by the angular declination below the horizon. *Nature* **414** (6860), pp.197–200.

Parker, A.R. (2011) On the origin of optics. *Optics & Laser Technology* **43** (2), 323–9.

Proffitt, D.R. (2006) Embodied perception and the economy of action. *Perspectives on Psychological Science* **1** (2), pp.110–22.

Proffitt, D.R. (2013) An embodied approach to perception: by what units are visual perceptions scaled? *Perspectives on Psychological Science* **8** (4), pp.474–83.

Proffitt, D.R. and Linkenauger, S.A. (2013) Perception viewed as a phenotypic expression. In: Prinz, M. Beisert, and A. Herwig (eds) *Action Science: Foundations of an Emerging Discipline*. MIT Press, Cambridge, MA, pp.171–98.

Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2005) Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience* **17** (6), pp.884–92.

Sedgwick, H. (1986) Space perception. In: Boff, K.L., Kaufman, L. and Thomas, J.P. (eds) *Handbook of Perception and Human Performance, Vol. 1: Sensory Processes and Perception*. Wiley, New York, NY, pp.128–67.

Shaffer, D.M. and Flint, M. (2011) Escalating slant: Increasing physiological potential does not reduce slant overestimates. *Psychological Science* **22** (2), pp.209–11.

Shaffer, D.M., McManama, E., Swank, C., and Durgin, F.H. (2013) Sugar and space? *Not the case: effects of low blood glucose on slant estimation are mediated by beliefs. i-Perception* **4** (3), pp.147–55.

Shapiro, L. (2013) When Is Cognition Embodied? In: Kriegel, U. (ed.) *Current Controversies In Philosophy of Mind*. Routledge, New York, NY.

Stefanucci, J.K. and Geuss, M.N. (2009) Big people, little world: the body influences size perception. *Perception* **38** (12), pp.1782–95.

Stefanucci, J.K. and Geuss, M.N. (2010) Duck! Scaling the height of a horizontal barrier to body height. *Attention, Perception & Psychophysics* **72** (5), pp.1338–49.

Wheatstone, C. (1838) On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London* **128**, pp.371–94.

Williams, L.E. and Bargh, J.A. (2008) Experiencing physical warmth promotes interpersonal warmth. *Science* **322** (5901), pp.606–7.

Wilson, A.D. and Golonka, S. (2013) Embodied cognition is not what you think it is. *Frontiers in Psychology*, **4**.

Wilson, M. (2002) Six views of embodied cognition. *Psychonomic Bulletin & Review* **9** (4), pp.625–36.

Witt, J.K. (2011) Action's effect on perception. *Current Directions in Psychological Science* **20** (3), pp.201–6.

Witt, J.K. and Proffitt, D.R. (2008) Action-specific influences on distance perception: a role for motor simulation. *Journal of Experimental Psychology: Human Perception and Performance* **34** (6), pp.1479–92.

Witt, J.K., Proffitt, D.R., and Epstein, W. (2005) Tool use affects perceived distance, but only when you intend to use it. *Journal of Experimental Psychology: Human Perception and Performance* **31** (5), pp.880–8.

Woods, A.J., Philbeck, J.W., and Danoff, J.V. (2009) The various perceptions of distance: an alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance* **35** (4), pp.1104–17.

Wraga, M. (1999) The role of eye height in perceiving affordances and object dimensions. *Perception & Psychophysics* **61** (3), pp.490–507.

Zhong, C.-B. and Liljenquist, K. (2006) Washing away your sins: Threatened morality and physical cleansing. *Science* **313**(5792), pp.1451–2.

Reply to Firestone

Chaz Firestone has produced a fine chapter. It presents my "moderate" approach to embodied cognition with great clarity; it shows appreciation for the novelty of the approach as compared with other proposals; and it sees the advance of this approach over its rivals. Firestone generously compliments it as a "refreshing and potentially game-changing approach." That was the good news, of course. In the rest of the chapter, Firestone focuses on a particular example explored in detail in Goldman (2012), a putative example of moderate embodiment in *perception.* His verdict on this example is thumbs down. It just isn't really an example of embodied cognition according to my own criteria. (A few explanatory remarks will follow shortly.)

What is my response? Firestone is correct; he has convinced me of his negative thesis. When properly interpreted, he argues, the phenomenon of "body-scaling" borrowed from Dennis Proffitt and colleagues turns out not to be an instance of cognition in the B-format sense. The remainder of this brief reply considers the consequences one might wish to draw from this concession.

The quickest and easiest move is one that Firestone himself anticipates: the proffered example of moderate embodied cognition simply doesn't live it up to its billing. Just forget it, then. What do I stand to lose? In (Goldman 2013) my central thesis about embodied cognition is formulated as follows: "a significant amount of human cognition has its origins in representations of one's own body" (p. 104). The phrase "a significant-amount" certainly leaves plenty of wiggle room. The satisfaction of such a criterion is by no means endangered if we choose to drop the original claim that perception (or much of it) is embodied. Innumerable other possible examples of embodiment are out there in possibility space. Quite a few of these were sketched in Goldman (2012); and even Firestone raises

Goldman and His Critics, First Edition. Edited by Brian P. McLaughlin and Hilary Kornblith.

no issue about them. So let's stop worrying; there's no reason to abandon our proposal or try to dilute it so as to please parties of every stripe.

Ending the discussing in this fashion, however, might be a bit abrupt and arguably sub-optimal. As Firestone points out, there is another alternative: tweak the criteria for embodiment so that body-based perceptual scaling *also* counts as moderately embodied. Let us explore this possibility.

First, let us review Firestone's reason for denying that body-based scaling in visual perception exemplifies the criteria for B-format based embodiment. Firestone explains that the information-processing that underlies body-based scaling in perception is entirely visual in nature, involving only geometric transformations on optical information. The visual system achieves its result by exploiting visual information *created by the body,* not by "redeploying" interoceptive bodily representations (for example, internally sensing one's height). Perception can and does represent space relative to the perceiver's body, but without employing B-formats. So, what kind of tweaking would accommodate visual perception as embodiment?

One obvious possibility would be to introduce a *disjunctive* criterion of embodiment: *either* embodiment in the B-formats sense *or* embodiment in the sense of one's body having a causal effect on a given type of cognition. The second disjunct was one of the earliest characterizations of embodiment that Vignemont and I considered in our joint paper (Goldman and de Vignemont 2009). Here was one formulation: there is embodied cognition when one's actions and other body-related traits (for example, posture) have an important causal role in cognition. The trouble is that really trivial cases would satisfy this criterion, thereby allowing cognitions to qualify as embodied far too readily. Nobody doubts that opening or closing one's eyes affects one's perceptions. Is this a reason to allow perception to qualify as embodied? Down this road, therefore, lies far too permissive a condition for embodiment. It would utterly trivialize the notion of embodiment cognition. Better, then, to stick with the B-format conceptions of embodiment as originally proposed, and not disjunctivize it with a second conception, even if it means abandoning the visual perception example proposed earlier (for Firestone's reason).

References

Goldman, A. (2012) A moderate approach to embodied cognitive science. *Review of Philosophy and Psychology* **3** (1), pp.71–88.

Goldman, A. (2013) The bodily formats approach to embodied cognition. In: Kriegel, U. (ed.) *Current Controversies In Philosophy of Mind*. Routledge, New York, NY. pp.91–108.

Goldman, A. and de Vignemont, F. (2009) Is social cognition embodied? *Trends in Cognitive Sciences* **13** (4), pp.154–9.