Interface Problems in the Explanation of Action
Daniel C. Burnston
Tulane University, Philosophy Department, and Tulane Brain Institute.

dburnsto@tulane.edu

Abstract

When doing mental ontology, we must ask how to individuate distinct categories of mental states, and then, given that individuation, ask how states from distinct categories interact. One promising proposal for how to individuate cognitive from sensorimotor states is in terms of their *representational form*. On these views, cognitive representations are propositional in structure, while sensorimotor representations have an internal structure that maps to the perceptual and kinematic dimensions involved in an action context. This way of thinking has resulted in worries about the interface between cognition and sensorimotor systems—i.e., about how representations of these distinct types might interact in performing actions. I claim that current solutions to the interface problem fail, because they have not sufficiently abandoned intuitions inspired by faculty psychology. In particular, current proposals seek to show how cognitive states can enforce prior decisions on sensorimotor systems. I argue that such "determination" views are the wrong kind of views to adopt, given the form distinction. Instead, I offer a proposal on which propositional representations can at best bias us towards certain kinds of action. This kind of view, I argue, appealingly distributes the explanation of action across distinctive contributions from cognitive and sensorimotor processing.

Acknowledgements

1. Introduction

In doing mental ontology, we must answer at least the following three questions. First: what are the categories of mental states and processes? Second: what are the appropriate definitions of the categories, and how do we individuate them? Third: how do states from the distinct categories interact to bring about mental and behavioral phenomena?

Traditionally, categories of mental states have been defined according to what I will call the "faculties approach." On this kind of view, for example, "cognition" underlies wholly different mental functions than "sensorimotor" systems. Cognition makes decisions, plans actions, and undertakes rational deliberation, with perception and motor processing being reduced to somewhat brute input and output devices. Many results in cognitive science have suggested that this kind of distinction is implausible—that perceptual and motor representations are involved in such processes as deductive reasoning, conceptual knowledge, and action planning (see, for example, Barsalou 1999, Hommel et al. 2001). In large part, these results promote the idea that sensorimotor representations are more *sophisticated* than traditionally thought. They can represent complex categorical structures and relations, can be employed in imagery, held in memory, etc.

One possibility, given these results, is to be revisionary about the first question, and to suggest that there simply are no distinct mental categories corresponding to the terms 'cognition', 'perception', and 'motor representation' (see, e.g., Spivey 2008). I will not engage with these views here. Rather, I will take up the possibility of being revisionary about the second question: namely, maintaining a distinction between cognition and sensorimotor systems by re-defining these categories. One way to do so is to abandon the idea of distinct faculties, , and instead to differentiate cognition from sensorimotor systems in terms of their representational *form* (Burnston 2016). The form distinction says nothing about the respective domains of cognition versus sensorimotor representation. Instead, it posits that cognitive representations are propositional and lexical in structure, whereas perceptual and motor representations have an internal structure that maps to the dimensions of perceived stimuli and motor movement, respectively. The hope is that this kind of distinction can be maintained even while granting the results that problematize the faculties approach.

At first, adopting the form distinction between cognition and sensorimotor systems seems much less revisionary than abandoning the distinction altogether. I will argue that, in fact, the form distinction has wide-ranging consequences for our understanding of mental architecture that have not been sufficiently explicated. I will show this by looking at the third question, and considering what kinds of relationships we need to posit between cognitive and sensorimotor representations if we adopt the form distinction. Several philosophers have noted the importance of this question, which has been dubbed the "interface problem" (Mylopoulos and Pacherie 2016, Butterfill and Sinigaglia 2014) Basically, the problem is this: producing actions seems to involve sophisticated semantic and causal interactions between cognitive and sensorimotor representations, but it is unclear how these interactions come about given that, on the form distinction, these representations are very different in structure.

I will argue that extant purported solutions to the interface problem fail, not primarily due to internal problems (although I will point out several), but because they frame the problem in a bad way. They posit relationships that make sense on faculties approaches, but are both implausible and, more importantly, *not the kind of relationships we should be looking for*, if we adopt the form distinction. In particular, extant proposed solutions seek to uphold the "causal

theory of action." On this view, practical reasoning—a cognitive process—culminates in a propositional intention to φ (e.g., to go to the store).[1] This intention causes specific motor representations to be tokened, and these implement the action. An action is intentional just in case the motor representation that implements it is caused in the right way by the intention to do it. I will use the term "determination" to refer to this kind of unidirectional causal and semantic relationship.[2]

Determination relations make a lot of sense on the faculties approach. If one thinks that sensorimotor states cannot be involved in decision-making or action planning, and that these are the sole purview of cognition, then any action that is planned or decided upon *must* involve some way of propagating the content of a cognitively arrived-upon decision to the motor representations that enact it (cf. Pacherie, 2000, p. 423). If one thinks that rationality has nothing to do with sensorimotor representations, then close semantic relationships between them and cognitive states are required to "rationalize them" (Mylopoulos & Pacherie, 2016, p. 5).

I will argue that determination is not the kind of relationship we should be seeking, given the form distinction. Thus, we shouldn't attempt to solve the interface problem as it has been posed. We should instead *dis*solve it by positing a different kind of relationship. Since, on the form distinction, cognitive states are fundamentally lexical in structure, I will assume that evidence about how lexical representations interact with sensorimotor ones is a good guide to thinking about the relevant relationships. I will draw on empirical and theoretical perspectives from cognitive linguistics and embodied cognition research to suggest that propositional representations can at best *bias* processing toward a *range* of possible sensorimotor outcomes. Biasing is a fundamentally different relationship than determination—one that admits that significant contributions towards action planning and decision must come from sensorimotor processing, outside of the strict commands of cognition. Indeed, once we have embraced a form distinction, this is exactly the kind of view we should want.

I'll proceed as follows. In section 2 I'll say more about what I mean by determination relations, and show how determination is at work in proposed solutions to the interface problem. I'll then argue that determination relations, of any sort, fail to account for what I call the diversity/specificity problem—that any propositional intention can correspond to a potentially wide range of sensorimotor representations. I think that this problem is unsolvable, but it's also only a problem for determination views. In section 3 I'll outline the biasing view and argue that it can make sense of a number of empirical results that determination views can't. Finally, in section 4, I'll propose a general view of the role of biasing relationships in action, which attributes *an* important causal role for propositional states, but abandons determination relations.

2. Determination Relations and Failed Solutions

---

[1] There are differences amongst causal theorists about whether intentions are reducible to belief-desire complexes, and whether all intentional actions *must* involve intentions (Davidson 1980). I will target the causal theory as expressed by theorists who think that intentions are distinct mental states, which stand in causal relations to the representations that guide and control the action—e.g., Bratman (1999), Searle (1983), and Pacherie (2000).

[2] Positing determination relations does not entail that there are *only* unidirectional causal relations between representations. Pacherie (2008), for instance, suggests that bottom-up interactions allow propositional intentions to monitor and correct downstream processing. This is compatible with the initial content-determination being top-down, and indeed, the interface problem is raised here again in the reverse direction. I thus won't engage further with this proposal.

The interface problem is basically one of coordination—the causal theory posits that the particular content of a given intention to φ (say, to fry eggs) needs to be coordinated with the content of the sensorimotor representations that implement the action that the intention has encoded. Current proposed solutions posit a *very* close coordination, wherein the intention causes a specific motor representation to be tokened. According to Butterfill and Sinigaglia (2014, henceforth, B&S), propositional intentions contain indexicals which can refer ostensively to sensorimotor representations. According to Mylopolous and Pacherie (2016, henceforth, M&P), propositional intentions contain no such demonstrative element; instead the content of the propositional intention is propagated into the motor system by a "content-preserving causal process" mediated by abstract motor schemas and attention.

I'll assume that the sensorimotor processes involved in action work roughly along the lines of the "Theory of Event Coding" (TEC) offered by Hommel and colleagues (Hommel et al. 2001, Memelink and Hommel 2013). On this view, the representations that implement and control action consist in bindings between perceptual representations of objects and motor representations of body movement. So, if I am making breakfast, I might first bind the motor representations involved in pouring water to a perceptual representation of my kettle, followed by binding the motor representations involved in frying eggs to perceptual ones of my egg carton, spatula, frying pan, etc. Typically, we must represent both the perceptual *starting* conditions and the perceptual *end* conditions of the action (e.g., an egg in the carton and an egg fried on the plate). The motor representation is the one that carries us from the starting to the ending conditions. We can create and store these bindings off-line, in what Hommel et al. refer to as "event files," and instantiate them when the time is right—action control consists in tokening the appropriate bindings in the appropriate circumstances. I'll discuss the structure of perceptual and motor representations in more detail below. The general question to be addressed is how propositional representations might interact with this kind of system.

Here is a general formulation of determination, meant to cover the multiple proposals.

- Determination: A determination relation is one in which the tokening of a state X with a content C, given background conditions, results via a unidirectional process of content-causation in the tokening of a second state Y.[3]

In a determination relation there is a unidirectional causal pathway from the tokening of X to the tokening of Y, and the reason why that process goes the way it does—i.e., the reason why Y is tokened rather than some other state—is because of the content of X. Hence, when we ask why Y occurred, the explanation is that X, with its specific content, occurred, and the appropriate causal path from X to Y also occurred. When we want to explain why a particular action controlled by motor representation Y occurred, we cite the content of intention X. This captures the idea that is central to the causal theory, namely that cognitive states are the "primary reason" for our actions. The interface problem, on determination views, is to see how a propositional intention can be the causal basis for actions implemented by representations in sensorimotor format.

Determination relations are compatible with other factors serving as intermediaries in the determination process. What makes the relationship one of determination is that these other factors are simply enablers—they facilitate the process by which the content of X results in the

---

[3] In the print version of the article, a mistake was introduced during copyediting, with 'via' being replaced by 'in'. This version is the appropriate wording.

tokening of Y. This loads the explanatory burden for the actions we perform (to the extent that they are intentional) on the content of the intention, as the causal theory does. I will argue, however, that this overall picture is wrong, given the form distinction. The gulf between the two kinds of content is not bridgeable by any determination relation.

Here is an example of determination in a neutral context. I recently had to purchase some eyeglasses in a foreign country after I'd lost mine at a conference. I was informed that there was a sale: two pairs for the price of one. After making my selections, I was informed that, despite the sale, the first pair alone was £175 and both pairs together were £215. The explanation was that the price of the lenses was greater when buying both pairs than when buying just one.

Let's suppose I was justified in being frustrated by this policy. What explains my frustration? Well, it seemed to me that the *content* of my decision, namely to get both pairs, *determined* the price of the glasses. The price, given my decision to get two pairs, was a specific amount, and this amount was distinct from what it would have been had I made the decision to get one pair. Given the background conditions, the explanation for the specifics of the price was the content of my decision. The tokening of my decision (state X) with its specific content (to get two pairs) determined the particular price (state Y) that was charged.

The background conditions were the store sale policies. Of course, these had to be in place before my decision could cause a specific change in the price. But they were not themselves the explanation of the price in this particular instance. *I* had to make the decision about how many pairs to purchase, and it was on the basis of the content of my decision that the price was determined. My frustration was due to the fact that, this determination relation being in place (and given the plausible mereological claim that lenses are a proper part of pairs), I literally could not make the decision to get two pairs for the cost of one.

In the action case, the background conditions are the nature of our sensorimotor systems—the kinds of things they can represent, their computational structure, etc. Obviously, if our perceptual systems could not represent cups, and our motor systems could not represent lifting movements, my decision to lift the cup couldn't eventuate in these representations being tokened. According to the determination intuition, sensorimotor systems can represent a range of objects upon which we can act, and a range of motor plans to act upon them. The *reason* why a particular object and motor plan are represented in a given instance is because we have the intention to act on the object for some particular goal—for example, we represent the motor plan of lifting with regards to the cup because we have the intention to take a drink. The goal of solutions to the interface problem is to show how the forming of an intention to φ results in the tokening of the particular sensorimotor representations that actually govern the action.

B&S start out by touting the sophistication of sensorimotor processing—they suggest that it can be employed in motor imagery separately from the actual act, and that this can occasionally be sufficient to ground the "outcome-directedness" of an action (2014, p. 120). The problem, according to them, is to show how intentions and motor representations get coordinated, such that their content comes to "match." Their story goes like this: when one is contemplating an action, one can engage in sensorimotor imagery, and thereby generate potential motor plans for interacting with objects. When one forms an intention, one decides to do *that*, where the indexical refers to a motor plan generated by one's sensorimotor systems. So, the tokening of the intention with its content (do 'that') results in a particular motor representation out of the ones represented in imagery in fact being tokened and controlling the action. Since the match between the content of the intention and that of the motor plan is due to the content of the intention (in this case, the indexical content), the indexical solution is one form of determination

view. So, despite the increased sophistication of sensorimotor processing on their view, B&S suggest that the indexical view is only a "refined and extended" version of the causal theory (2014, p. 120).

M&P's solution eschews indexicality in favor of a hierarchy of decreasingly abstract representations which propagate the content of the propositional state into the motor system. The idea here is that there are intermediary processes that take the propositional content and implement it in the specific situation faced by the actor. First, we have future-directed intentions, which specify the action prior to the situation in which it is to be performed. These give way to "proximal" intentions, which are intentions to act now, in the current situation. Both of these are propositional. The connecting device that is supposed to take the propositional content into the motor system is a "motor schema," which is a highly abstract, but still motoric in format, representation. According to M&P, a motor concept like 'grasp' is associated with an extremely general motor program, involving something like reaching out the arm. In action generation, there is first a propositional intention, which activates a highly abstract motor representation, and then this representation is enriched with detail to fit the current circumstances. Since this process is "content-preserving," and starts with the formation of an intention, and since the specific motor representations tokened are the result of the process, the view posits a determination relation.

Neither of these views succeed at establishing determination, I claim, because they both fall prey to what I will call the "diversity/specificity problem." The issue is that any one propositional representation might correspond to a *range* of *particular* motor representations. (I use 'correspond' here intentionally vaguely, to refer to semantic coherence between a proposition and a motor representation without presupposing how that relation is implemented.) But if there are multiple motor representations with *different* specific contents, which correspond to the same intention, then the content of the intention can't determine between those contents. At first, this problem sounds like detail-mongering, but I think it is actually quite deep, and points up the tensions in maintaining the causal theory along with the form distinction.

Think about the lexical concept 'dog'. Perceptually speaking, 'dog' corresponds to a *very* large number of possible perceptual representations, where these differences include differences in subtypes and, even with a subtype, differences in their precise perceptual properties. 'Dog' corresponds both to Dobermans and Pekingese, but even within each subcategory there are a wide range of sizes, colors, postures, fur patterns, etc., that correspond to the concept. Motor concepts are similar. 'Kick' corresponds to roundhouse kicks, sweep kicks, nudges with the foot, etc. Even within each of these different types, motor kinematics will vary depending on the properties and positions of the objects being kicked. So, even once I've decided to kick someone, I still need to select a particular subtype of that action *and* execute the kick appropriately depending on the properties of the target (how tall or short they are, their current posture, etc.). These parameters of the action *must* be specified by the motor representation for the action to be carried out. The worry is that, given that propositional content doesn't have the same structure as motor representations, it doesn't have the resources to determine these particulars.

The indexical solution fails to account for this combination of diversity and specificity. According to B&S, a particular motor representation is indexed by the demonstrative element in the intention. The decision to index the motor representation is still due to the cognitive/propositional system—tokening the particular motor representation is supposed to be the *result* of the intention "selecting" that motor representation via denotative reference.

However, given a range of possible particular motor representations corresponding to the intention to φ, there are no resources to explain how that reference is supposed to be fixed. Cognition itself doesn't have access to the particulars of the motor content—it is purely propositional on the form distinction, after all. But it is precisely this content that has to be selected in order to settle on one motor representation rather than another. M&P thus, quite rightly, complain that B&S's solution presumes a solution to the interface problem rather than providing it.

The motor schema solution proposed by M&P, however, fares no better. The worry can be phrased in terms of a dilemma. Suppose that the motor schema associated with the lexical concept is *very* abstract, such that it generalizes over all of the particular motor representations that might correspond to it. In this case, the schema itself will fail to account for specificity, since it will run out of content to contribute long before any actual motor representation is tokened. If the schema is less abstract, however, such that it directly corresponds to some subset of the diverse motor representations, then it fails to account for diversity, since it will only be able to determine a subset of the motor representations corresponding to the lexical concept.[4]

This discussion allows us to address some intuitive responses. The first is to meet with the worry about specificity by simply adding more specific lexical concepts to the presumed intention—for example, to 'sweep kick' rather than simply 'kick'. As we saw above, however, simply settling on a subtype won't solve the problem, since the particular motor representation instantiated will need to be modified depending on the circumstances of the act. A corresponding move, which is sometimes made in philosophy of action (e.g., Bratman 1999, ch. 3), is to posit epicycles of practical reasoning and more fine-grained intentions to solve the problem. Maybe my intention to kick Fred gets me into his vicinity, then I engage in more practical reasoning to determine precisely that I want to kick him thusly. On some views, I now token a new kind of intention (what M&P call a P-intention), which instructs my motor system precisely how to act. Given the above, however, it is clear that these proposed solutions only push the problem down the road. If these intermediate processes involve something like M&P's schema solution, then they face the problem of insufficient contents. If eventually they terminate in an indexical, then they run into the problem for B&S's solution: the propositional state will lack the needed content to refer to a specific motor representation.

The same problem applies to another standard idea, which is that we have a hierarchy of more abstract to more specific representations, with the ones at the top encoding the abstract intention (or motor schema) and the ones at the bottom encoding motor specifics. No matter how many levels in the hierarchy one posits, the diversity/specificity problem will inevitably arise between the intervening levels. The abstract content at the top, by definition, just doesn't *have* the specific content. So positing more, littler steps of content determination is just re-raising the problem on smaller and smaller scales.

A final common proposal is to introduce some new process, which is not itself an intention, to implement the needed transition from the propositional to the motor. M&P in fact make this move, invoking *attention* as the process that links the abstract motor schema with specific parameters of objects such that our motor representations can target them (cf. Wu 2011).

---

[4] M&P occasionally talk in terms of "prototype" actions encoded in a motor schema. B&S raise the possibility that motor representations encode the "invariant" aspects of action. However, positing prototypes or invariants faces the same problem as positing highly abstract schemas. Given that we act on particulars, not prototypes, there will always be a gap between the content of the prototype/invariant and the particular representations tokened—a gap which determination views have no resources to fill.

However, so long as it is the content of the intention that attention is supposed to implement (M&P state, vaguely, that attention is "often guided by intention"; 2016, p. 23), this is no solution at all, because it is left entirely mysterious how attention takes the content of the intention and interfaces it with more specific motor representations. Again, a solution has been presupposed rather than offered.

Given all of these difficulties for determination views, it's worth taking a step back and asking whether the interface problem—construed in terms of determination—is really one we should be trying to solve. Suppose this is my first visit to your home, and I am helping to set the table before your dinner party. You ask me to bring you the soup ladle on the counter in the kitchen, and (being a reasonable guest) I determine to do precisely what you ask. Given that I haven't seen the ladle before, the intention to "bring the ladle" is very helpful, and *precisely for the reason* that I don't know what it looks like, its position on the counter, etc. Ideally, the intention should help me get the ladle no matter what the particulars of the scene turn out to be. In order to serve this role, it seems like we should very much *not* want our intentions to determine the content of our sensorimotor states. On the other hand, if I have been to dozens of your dinner parties, and I *do* know exactly what your ladle looks like, where it always resides on the counter, etc., then it seems I have no need of the intention to determine my sensorimotor representation. I just need something to help me orient my actions towards the ladle as opposed to something else. Determination is, one might reasonably suppose, a stronger relation than is needed for this kind of role.

I think that, absent theoretical affiliations inspired by the faculties approach, this description of the role of intentions (vaguely stated as it is) is perfectly intuitive. Assuming we can flesh out in a conceptually and empirically satisfying way what a non-determining role for intentions looks like, we should be open to it. We should dissolve, not solve, the interface problem.

3. A Dimensional Biasing Account

If we wish to maintain the form distinction, we need a view of how propositional and sensorimotor states interact that strikes the appropriate balance between diversity and specificity. Fortunately, just such an approach is available: in this section I will argue that propositional states *bias sensorimotor dimensions*, rather than determining the contents of motor representations. This results in a view on which intentions can *potentiate* a *range* of possible sensorimotor representations, but don't determine which is enacted.

The first step is to recognize that perceptual and motor representations in fact potentiate each other automatically (i.e., outside of control by intention), and this is a significant component of the generation of an action. The second step is to recognize that propositional states *do* interact with these processes, but that this relationship is best construed in terms of biasing, not determination. I will now walk through some evidence supporting each of these steps.

3.1. Dimensions and Automatic Processing

Perceptual and motor representations consist in represented values along relevant dimensions. When I refer to "sensorimotor" dimensions in general, I mean both the kinematic dimensions of an action and the perceptual dimensions towards which acts are oriented. Types of perceptual objects are higher-order, holistic representations that relate a variety of subsidiary dimensions

(Folstein, Gauthier, and Palmeri 2012, Goldstone and Hendrickson 2010). For instance, male faces, compared to female faces, tend to have larger noses, more pronounced chins, more space between the nose and upper lip, etc. For other categories, such as 'cup' or 'frying pan', the dimensional structure is simpler. Cups characteristically vary on only some dimensions (volume and shape dimensions), while frying pans have a quite characteristic set of values along shape dimensions (one flat bit and one extended bit). Recognizing an object perceptually involves perceiving it as inhabiting a "space" that is definitive of the type of object it is, along the relevant dimensions (cf. Burnston, 2016).

A motor action such as grasping or kicking consists in a higher-order relationship between bodily dimensions. For instance, grasping involves a relatively quick extension of the elbow followed by a closing of the hand. A power grip and a precision grip involve different higher-order relationships. Of course, as discussed above, there are ranges along these dimensions. A grasped cup can have a variety of specific perceptual properties which must—if the grasping is to be successful—be bound to particular values along motor dimensions. Moreover, with regards to cups, *grasping to drink* will involve a different motor representation than *grasping to place*. While both will involve a sideways rotation of the wrist and a contraction of the fingers, the former will involve an extension of the elbow followed by a contraction, and the latter will involve an extension followed by a further (or at least differently targeted) extension. Which of these programs is enacted, along with the particular values of the assorted extensions, contractions, etc., depends upon the visualized end-state of the act.

We are now in a position to flesh out the idea behind the TEC a little further: an action representation is the binding of certain values along perceptual and motor dimensions. The particular values of the dimensions defining (say) a precision grip must be bound to specific dimensional values for the object to be gripped in that manner. Importantly, there is evidence that a significant factor in this binding is bi-directional automatic association between perceptual and motor dimensions.

Here are a few examples. First, in the perceptual-to-motor direction: if one is asked to grasp an object with a particular grip type (e.g., precision versus power grasp), performance is slower if the required grip is not the standard one used to grasp that object (Tucker and Ellis 2001). This suggests that the normal grip is automatically activated by perception of the object. Similarly, perceiving an object that one normally moves *away* from the body influences actions differently than one that moves *towards* the body. If, in an experiment, a subject must respond to a stimulus by pulling a lever towards them, they will do so more slowly if the stimulus depicts an object that is more commonly moved away (Glenberg and Kaschak 2002). Lastly, photographs of the *effects* of a movement—that is, the expected perceptual consequences of the required movement—facilitate it (Elsner and Hommel 2001). In the motor-to-perception direction, just viewing a picture of a grasping hand facilitates identification of objects which could be gripped by a grasp of that type, and this finding is enhanced if the sizes match (Vainio et al. 2008). Similarly, planning a grasping movement primes subjects to differentiate items based on their size—size being a relevant dimension for a grasping act (Fagioli, Hommel, and Schubotz 2007).

This suggests that one factor involved in binding perceptual and motor dimensions in actions is automatic association between certain perceptual and certain motor dimensions. The idea behind the biasing view is that there are certain dimensions characteristic both of motor movements and perceptual objects, and that each of these contains a range of possible values. Tokening a lexical concept as part of an intention can bias the motor system to act *somewhere* along these dimensions, without determining their precise values (Memelink & Hommel, 2013),

and thus *potentiate* certain *classes* of effects. Vitally, this influence is inherently probabilistic, rather than deterministic. Tokening lexical representations makes certain classes of sensorimotor representations more likely—they are tokened more easily, quickly, etc. While these interactions might be *a* causal factor in bringing about actions, they are not a determining factor. Instead, they interact with the kinds of bindings already being undertaken by perceptual and motor systems, and the combined interactions result in an event file, with no one part of the process determining the outcome. This view, I contend, accounts for a range of effects that are simply hard to square with determination views.[5]

3.2. Results in Favor of a Biasing View.

If you look only at individual cases of lexical-motor interaction, they can look like determination relations. In an attentional paradigm, being told to look for a green object facilitates pop-out of that kind of stimulus from a background of distractors. Being told to find a green *square* produces similar results (Weidner and Muller 2009). If we're told to grasp an object in front of us using a precision grip, we do it. So what's the problem? It turns out that reading these intuitive results as determination relations misses much of how the mechanisms actually work. Here are some results that can easily be accounted for with dimensional biasing, but not with determination relations.

3.2.1. Diversity and Categorical Overflow.

As we saw in the previous section, a problem for determination views is the diversity of possible sensorimotor representations that might correspond to a given lexical concept. It turns out that this problem is empirically validated—multiple distinct seonsorimotor representations *are* in fact potentiated by lexical ones. This diversity comes in two flavors: in one, a single lexical concept can potentiate a range of possible dimensions and dimensional values; in another, a single potentiated dimension can influence multiple distinct categories, a property which I have elsewhere called "categorical overflow" (Burnston, 2016).

Categorical overflow can be seen in motor priming. Myung et al. (Myung, Blumstein, and Sedivy 2006) showed that if subjects are primed with 'piano', their eye movements are drawn by pictures of typewriters. The authors interpret this as due to the similar motor representations tokened while acting with the two objects, despite their relatively low level of either functional or visual similarity. Similarly, viewing a photograph of a dustpan primes recognizing an item in a subsequently presented photograph as a spatula (Helbig, Graf, and Kiefer 2006). The idea here is similar: since one makes similar arm movements with dustpans and spatulas, activating these dimensions facilitates both despite their categorical differences.

On the perceptual side of the ledger, both diversity and categorical overflow are evident in studies of attention. If one primes a subject with the word 'snake' while presenting a display of objects, subjects will automatically focus attention on any object that is *either* elongated and

---

[5] As noted in the introduction, I am making the assumption here that, since natural language and propositional states have a similar structure, we can learn about the kinds of interactions undergone by the latter by studying the former. Defenders of determination views are free to attempt to define a difference between lexical and propositional representations but, given the definition of cognitive representations in terms of form, this seems a difficult strategy. I won't say more about it here; hopefully the positive appeal of the biasing view will be sufficient to obviate this kind of move.

curvy *or* coiled (Huettig and Altmann 2007). Importantly, this effect occurs even when the object is *not* in fact a snake (e.g., a coiled cable). This shows (i) that 'snake' doesn't determine particular values along shape dimensions, but instead potentiates a variety of possible values along them, and (ii) that what is important in the relationship between the lexical concept and perceptual processing is not that it says something *specific* that's definitive of snakes, but that it primes a set of dimensional values that can perhaps be shared in common between snakes and other kinds of things. This flexibility is what allows the biasing view to account for diverse effects across trials, but specific effects within each. And, as we saw above, determination views fail to account for these properties.

3.2.2. Importance of Context.

If you instruct a subject in an experiment to grasp a wooden block in front of them, they'll do so. But this effect is modifiable by presenting a lexical stimulus. Glover et al. (2004) had subjects grasp a block, but before they did they primed them with auditory presentation of a word. Intriguingly, subjects' initial grip aperture was larger or smaller depending on the word that was presented. Presentation of 'apple' resulted in an initial grip aperture that was larger, for instance, than when 'grape' was presented. However, labels didn't prime subject to make *precisely* the grasp motion that would be involved in grasping either grapes or apples—indeed, the precise aperture of the grip given a particular word differed depending on the size of the actual block they were about to grasp. It's unclear what determination views would even say is going on here, but dimensional biasing has an easy account. There is a certain dimensional value for aperture defined by the object that one is attempting to grasp, and presentation of lexical labels shifts that value somewhat closer toward the range of values associated with the primed concept. The result is an interaction effect between primed dimensional values and perceived ones.

On the perceptual side, there is evidence that perceptual categorization is heavily dependent on the actual perceptual context. For instance Sedivy et al. (1999) showed that when subjects are asked to look for a *tall* glass on a screen, they are quicker to focus attention on a stereotypically tall glass than to a shorter one. However, when a contrasting image is also present (i.e., an even smaller glass), even atypical 'tall' glasses attract attention much more quickly. Similarly, if subjects are asked to "pick up the blue comb," they will attend to a blue comb more quickly if there is a salient contrast item (e.g., a pink comb) in the display. These results are interpreted as suggesting that adjectives are interpreted contrastively—in other words, the lexical label only specifies a range of values along a dimension, whose particular values can be fixed with help from further information (e.g., contrast information) in the environment.

Finally, there is evidence that perceptual representations rely implicitly on contextual information that is simply not encoded in the content of the propositional intentions subjects have (Zang et al. 2016). If you have subjects focus search for particular orientations of shapes in a visual discrimination task, and the shapes occur at different points around a semantically unrelated and task-irrelevant object (such as a rectangle), after a brief time subjects will have a harder time locating the shapes when the object is removed. In this case, the surrounding perceptual context affects the perceptual processing of a named object in a way that is semantically unrelated both to the concept that is intentionally being searched for and the kinds of responses they intend to make.

Context effects like this make little sense on determination views, on which the content of the intention is responsible for which sensorimotor representations we token. Instead, these

results suggest that particular tokened values along sensorimotor dimensions rely vitally on bottom-up contextual information that is not encoded by the intention. A biasing view can account for how top-down influence might interact with these processes.

3.2.3. Facilitation versus Interruption Effects

On determination views, the content of the intention determines the content of the motor representation that controls the action. One would expect, on this view, that intentions generally facilitate actions. However, lexical labels can either facilitate or interrupt actions, depending on the context of the action, *even if* the action being performed corresponds to the meaning of the lexical concept.

To take just one example: it turns out that the timing of lexical concept activation matters a lot for whether it facilitates an action or interrupts it. If verbal labels connoting an action are presented *before* the action is undertaken, then they facilitate the action. However, if they are presented *during* the action, they in fact interrupt it (Boulenger et al. 2006). Here's an interpretation of these effects: prior to the adoption of a specific set of values for motor dimensions, the label biases a range of values and hence, since the enacted values are part of the set, facilitates those particular values. However, after a particular set of values has already been adopted, re-biasing *other* values, even within the same dimension, could in fact interrupt the action in progress. But if lexical influence works this way, then it doesn't determine actions.

3.2.4. Inter-trial Effects.

Finally, a range of studies show that there is *inertia* to represented dimensions. This is shown through switch costs that are involved in changing from representing along one dimension to another.

Return to attentional pop-out studies, in which a subject must detect a target item amongst distractors (Weidner & Müller, 2009). If, across two trials, the searched-for object is defined in different dimensions—for instance, if subjects first have to look for a green object, followed by looking for one that is pointing left—they are much slower to recognize the object in the second trial. For changing features *within* a dimension, however, the results are much different, either entailing no cost (when compared to a neutral condition in the first trial) or conferring a benefit. In color, for instance, there is a short switch cost if subjects must change from looking for a red object to a green one, compared to simply being asked to look for a green one. Still, however, these responses are slightly *quicker* than if the instructions in the first trial don't name a color. In orientation there is no cost al all in switching from, say looking for a leftward-oriented bar to a rightward-looking one, and having just looked for a rightward-oriented bar produces significant facilitation effects when you look for a leftward-oriented one, compared to a neutral initial trial.

Attention researchers interpret these results as suggesting that, even in feature-cueing (e.g., "look for green") circumstances, it is primarily dimensions that get primed, such that any feature along that dimension will be facilitated. But if it is whole dimensions that are primed, then lexical representations do not determine specific values.[6] Moreover, the fact that inter-trial

---

[6] Now, some have been tempted to say that only dimensions, and not features are potentiated (e.g., Memelink & Hommel, 2013), but this seems overly strong. At least in the case of color, there are specific advantages to tokening 'green' as opposed to simply tokening 'color' (although these advantages don't seem to hold for orientation). But

costs occur suggests a *limit* to the power of lexical representation to influence the dimensions that are active. Similar results occur in motor dimensions—once you have bound a motor representation in an event file, it is harder to use that same representation for a different event file (Hommel 2004). Together, these results show that lexical representations are limited in the specificity and efficacy of their influence on sensorimotor ones.

3.2.5. Upshot

Determination views posit a straightforward relationship between the content of an intention and that of the sensorimotor representations that govern particular actions, and exhaust themselves trying to show how that straightforward relationship is implemented. But the above results suggest that the relationship simply *isn't* straightforward. Instead, the effects are highly general, highly negotiable, and highly context-sensitive. The biasing view has the flexibility to account for this range of results. Diversity is due to the wide range of potentiated dimensional values; specificity is due to the interaction of these biasing effects with the bottom-up construction of particular representations to fit particular contexts. The biasing view respects the empirical facts of diversity and specificity, rather than trying to theorize them away.

 As mentioned above, biasing relations can look like determination relations if you squint right—if you hold everything else fixed in choosing action targets and motor representations, and you look at only one instance. All Buridan's Ass needs is a slight nudge to the left. But the nudge is simply that. On the biasing view, the nudge can be a big help without having to determine how the ass walks to the hay bale. I'll return to this kind of point in the final section.

3.3. Objections

One possible avenue for defending determination views from these kinds of considerations is to suggest that, in a particular intention, we combine a bunch of lexical concepts. If one concept can't determine specific content, then maybe multiple concepts can.

 Alas, this is not a very helpful option. If, say, 'grasp' doesn't determine the precise dimensions of motor representations, and 'grape' doesn't determine the particular perceptual properties of the objects to be grasped, putting them together doesn't do so either. There are still a range of grasp-types and still a range of grape-particularities that could go into the action—I might need to grasp a grape in many different motor (a grape lying alone versus a grape in the bunch) or perceptual (red versus green, bunched versus un-bunched) circumstances. At best we would now have a larger set of dimensions on the table, and the problem of putting the right perceptual values together with the right motor ones remains. The problem is precisely a problem of format. Adding more propositional content doesn't help if propositional content isn't the right kind of representation to perform a determination role.

 This is especially clear in the case of *novel* actions—actions that have not been performed before, like scratching your back with a floppy disk. According to Glenberg et al.'s (2009) "affordance meshing" hypothesis, lexical representations do two things. They facilitate

---

even admitting *some* feature-specific effects is not really much help in the overall discussion. 'Green', of course, is a determinable, and what counts as green for subjects can depend on the relevant contrast in the environment (Hardin 2008). An alternative, and as far as I can see unproblematic account, is to interpret the tokening of 'green' as potentiating a dimension *relatively*—namely, biasing certain values slightly more than others, where the details are determined by the interaction of this relative biasing with the actual information available in the environment.

identification of items to be acted upon in the environment, and they specify very vague roles in which objects might be related to each other. So 'scratch' doesn't say anything about floppy disks or backs, but does specify that whatever the implement and target of the scratching are, they must stand in certain roles with regards to each other (the implement moves across the surface of the target). Once the objects are indexed and the roles specified, one must then "mesh" the affordances of the particular objects. Given a particular floppy disk you've identified, you can then figure out a way to scratch your back with it. But the lexical instruction provides no content to say how this meshing should go. At this point, sensorimotor processing must take over and determine how to best perform the action. Further grammatical categories might be imported, but they simply add more abstract structure. For instance, presentation of the word 'into' automatically primes subjects to look at items that could serve as containers, before any further information (e.g., about the properties of the container) is mentioned (Chambers et al. 2002). Specifying abstract relationships, however, is not the same as determining content.

A final objection must be considered. I have talked a lot about biasing, reaction times, attention, and so on. But practical reasoning is supposed to be about *decision-making*. I intend to act thusly because I have considered my reasons and have *decided* to do it. And decision is a cognitive process, not a sensorimotor one.

I hope it will be obvious at this point that this is a faculties-based, rather than a form-based claim. As such, it is inadmissible as an objection.[7] Once the form distinction is admitted, one cannot simply stipulate that a particular old faculty-based concept must fall on one side or the other of the sensorimotor/cognitive divide. To see this, consider that there are models on which the decision to act needn't be determined by a prior propositional state. According to Cisek and Kalaska's (2010) "affordance competition" model, for instance, objects present us with a range of action possibilities, and these take on relative salience given our current motivations. So, if one is thirsty, then the affordance of a cup for bringing-to-the-mouth is simply more salient to you than the grasping-to-place affordance (or the prop-up-the-window-affordance, etc., etc.). Per Cisek, the event file set that is arrived at is the most salient "winner" of a "competition" between all of the action-affordances of all of the objects in one's environment. On this model, decision doesn't *require* prior propositional reasoning and the subsequent tokening of a propositional intention. The dimensional biasing view further explains how the tokening of a propositional state could have *a* role in regards to this process, but along with abandoning the necessity of a prior propositional decision, abandons the need for determination relationships to occur.[8]

I suspect that proponents of solutions to the interface problem will be inclined to try to back away from determination claims, and hence suggest that what I've proposed here is in fact an amenable filling-in of their accounts. But I don't think that's sustainable. Rather than try to anticipate details, I'll conclude this paper with a gauntlet—a general description of the role of intention that should be anathema to those in favor of the traditional causal view, but is highly compatible with the dimensional-biasing account. Those in favor of more traditional views must explain why the data supports the old model rather than this one.

---

[7] Indeed, when one asks subjects to perform a task that is incongruent with the affordances of an object (for instance, in the effects cited in section 3.1), error rates generally go up, suggesting, again, that the resources of the intention we "decide on" aren't sufficient to determine the action.

[8] I believe this perspective can be scaled up if combined with recent philosophical approaches that focus on *imagination* (Nanay 2016) and *skill* (Railton 2009) in decision-making, and that this is a necessary project to undertake if we adopt the form distinction. I won't pursue it further here, however.

4. The Grocery List Model of Intention

Imagine you have a grocery list.[9] You've thought it over, and you've decided what you "intend" to get. Imagine the same list in two distinct situations. In the first, you go to your normal grocery store, where you shop weekly. In the second, you are in a new town on the other side of the country, with slightly different brands, etc. By stipulation, the content of your list is the same in both circumstances. But it is obvious that the details of your actions, both in terms of motor and perceptual contents, will differ considerably. You may get a different brand of pasta sauce—with a red label, say, rather than a green one, or in a differently shaped jar. You may have to get it from the bottom shelf rather than the top one. Moreover, your overall strategy will differ. In your normal store, reading the 'sauce' item on the list will, along with the known perceptual context you find yourself in, facilitate the recall of a stored representation of the layout of the store, the surrounding perceptual circumstances (what else is on the same and nearby shelves), etc., and you will orient yourself to walk towards the pasta aisle. In the new store, you may adopt a very different event file. You may adopt a scanning behavior, walking up and down the aisles for perceptual clues about where the sauce might be (e.g., reading the signs above the aisles, looking down the aisle for items standardly grouped with sauce). Eventually you'll end up with sauce, but the sensorimotor representations underlying your success will differ considerably.

As above, if the *same* propositional content can have very *different* effects depending on the context, then it doesn't determine the content of sensorimotor representations. This situation is simply a generalization of the kinds of results I've been discussing, and the lesson is the same. I thus suggest that the role of propositional intentions is a lot like that of the items on a grocery list. We may "adopt" them, in the sense of holding them in mind or writing them down, and reciting them might bias us towards certain behaviors. But they don't determine what we do. Much of that work is due to the rough-and-tumble processing of sensorimotor representations in occurrent contexts.

Why is this role for intentions important? Even in your habitual store, you are beset by stimuli that afford actions, and that have varying degrees of salience, and these pull at you and can interrupt your plans. Even when I have adopted the goal of obtaining vegetables, I admit that I do often find myself drawn to the beer and ice cream aisles. The list biases you in favor of certain items rather than distractor items, such that you are more likely to attend to them and orient your actions towards them. But alas, this is only probabilistic. Sometimes I forget things on my list. Sometimes (often) I end up with more beer and ice cream than vegetables. In the new store, you don't already have a stored event file. You have to construct one despite these many distractions. The list (/intention) biases you towards certain *sets* of actions, such that, within your current context, you can try to figure out the best one. This "figuring out," on the view I've been trying to articulate, is itself a sensorimotor process.

I'd like to end by noting that, if we're already impressed enough by the sophistication of sensorimotor processing to adopt the form distinction in the first place, then this is exactly the kind of view we should want—one that distributes responsibility for more traditionally "cognitive" function across interactions between distinct representations. It's true, we may need to rethink some cherished concepts in these terms, including 'decision' and maybe even (gasp!)

---

[9] Anscombe (2000) used a grocery list example as an analogy for describing *when* an act is intentional—one intends to get butter if ending up with margarine is a failure of the action, not the list. My point is different: I think the role of the list is an appropriate *model* for the role of intention.

'rationality'. But we're already halfway down that road, and the sciences are offering up rich frameworks for thinking about these issues, such as the one I have tried to articulate here. We shouldn't lack the courage of our convictions.

REFERENCES:

Anscombe, Gertrude Elizabeth Margaret. 2000. *Intention*: Harvard University Press.

Barsalou, Lawrence W. 1999. "Perceptions of perceptual symbols." *Behavioral and brain sciences* no. 22 (04):637-660.

Boulenger, Véronique, Alice C. Roy, Yves Paulignan, Viviane Deprez, Marc Jeannerod, and Tatjana a Nazir. 2006. "Cross-talk between language processes and overt motor behavior in the first 200 msec of processing." *Journal of cognitive neuroscience* no. 18 (10):1607-15. doi: 10.1162/jocn.2006.18.10.1607.

Bratman, Michael. 1999. *Intention, plans, and practical reason*. 2 ed. Cambridge: Cambridge University Press.

Burnston, Daniel C. 2016. "Cognitive penetration and the cognition–perception interface." *Synthese*:1-24.

Butterfill, Stephen A, and Corrado Sinigaglia. 2014. "Intention and motor representation in purposive action." *Philosophy and Phenomenological Research* no. 88 (1):119-145.

Chambers, Craig G, Michael K Tanenhaus, Kathleen M Eberhard, Hana Filip, and Greg N Carlson. 2002. "Circumscribing referential domains during real-time language comprehension." *Journal of memory and language* no. 47 (1):30-49.

Cisek, P., and J. F. Kalaska. 2010. "Neural mechanisms for interacting with a world full of action choices." *Annu Rev Neurosci* no. 33:269-98. doi: 10.1146/annurev.neuro.051508.135409.

Davidson, Donald. 1980. *Essays on actions and events: Philosophical essays*. Oxford: Oxford University Press

Elsner, B., and B. Hommel. 2001. "Effect anticipation and action control." *Journal of Experimental Psychology: Human Perception and Performance* no. 27 (1):229-229.

Fagioli, Sabrina, Bernhard Hommel, and Ricarda Ines Schubotz. 2007. "Intentional control of attention: Action planning primes action-related stimulus dimensions." *Psychological research* no. 71 (1):22-29.

Folstein, Jonathan R., Isabel Gauthier, and Thomas J Palmeri. 2012. "How category learning affects object, representations: Not all morph spaces stretch alike." *Journal of Experimental, Learning, Memory & Cognition* no. 38 (4):807-20.

Glenberg, Arthur M, Raymond Becker, Susann Klötzer, Lidia Kolanko, Silvana Müller, and Mike Rinck. 2009. "Episodic affordances contribute to language comprehension." *Language and Cognition* no. 1 (1):113-135.

Glenberg, Arthur M, and Michael P Kaschak. 2002. "Grounding language in action." *Psychonomic bulletin & review* no. 9 (3):558-565.

Glover, Scott, David A Rosenbaum, Jeremy Graham, and Peter Dixon. 2004. "Grasping the meaning of words." *Experimental Brain Research* no. 154 (1):103-108.

Goldstone, Robert L., and Andrew T. Hendrickson. 2010. "Categorical perception." *Wiley Interdisciplinary Reviews: Cognitive Science* no. 1 (1):69-78. doi: 10.1002/wcs.26.

Hardin, CL. 2008. "Color Qualities and the Physical World." In *The case for qualia*, edited by Edmond Wright, 143-154. Cambridge, MA: MIT Press.

Helbig, Hannah Barbara, Markus Graf, and Markus Kiefer. 2006. "The role of action representations in visual object recognition." *Experimental Brain Research* no. 174 (2):221-228.

Hommel, B. 2004. "Event files: Feature binding in and across perception and action." *Trends in cognitive sciences* no. 8 (11):494-500.

Hommel, B., J. Musseler, G. Aschersleben, and W. Prinz. 2001. "The theory of event coding (TEC): A framework for perception and action planning." *Behavioral and brain sciences* no. 24 (5):849-877.

Huettig, Falk, and Gerry TM Altmann. 2007. "Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness." *Visual Cognition* no. 15 (8):985-1018.

Memelink, Jiska, and Bernhard Hommel. 2013. "Intentional weighting: a basic principle in cognitive control." *Psychological Research* no. 77 (3):249-259.

Mylopoulos, Myrto, and Elisabeth Pacherie. 2016. "Intentions and motor representations: The interface challenge." *Review of Philosophy and Psychology*:1-20.

Myung, Jong-yoon, Sheila E Blumstein, and Julie C Sedivy. 2006. "Playing on the typewriter, typing on the piano: manipulation knowledge of objects." *Cognition* no. 98 (3):223-243.

Nanay, Bence. 2016. "The Role of Imagination in Decision-Making." *Mind & Language* no. 31 (1):127-143.

Pacherie, Elisabeth. 2000. "The Content of Intentions." no. 15 (4):400-432.

Pacherie, Elisabeth. 2008. "The phenomenology of action: A conceptual framework." *Cognition* no. 107 (1):179-217.

Railton, Peter. 2009. "Practical competence and fluent agency." *Reasons for action*:81-115.

Searle, John R. 1983. *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Sedivy, Julie C, Michael K Tanenhaus, Craig G Chambers, and Gregory N Carlson. 1999. "Achieving incremental semantic interpretation through contextual representation." *Cognition* no. 71 (2):109-147.

Spivey, Michael. 2008. *The continuity of mind*. Vol. 40: Oxford University Press, USA.

Tucker, Mike, and Rob Ellis. 2001. "The potentiation of grasp types during visual object categorization." *Visual cognition* no. 8 (6):769-800.

Vainio, Lari, Ed Symes, Rob Ellis, Mike Tucker, and Giovanni Ottoboni. 2008. "On the relations between action planning, object identification, and motor representations of observed actions and objects." *Cognition* no. 108 (2):444-465.

Weidner, R., and H. J. Muller. 2009. "Dimensional weighting of primary and secondary target-defining dimensions in visual search for singleton conjunction targets." *Psychol Res* no. 73 (2):198-211. doi: 10.1007/s00426-008-0208-9.

Wu, W. 2011. "Confronting Many-Many Problems: Attention and Agentive Control." *Noûs* no. 45 (1):50-76.

Zang, X., T. Geyer, L. Assumpcao, H. J. Muller, and Z. Shi. 2016. "From Foreground to Background: How Task-Neutral Context Influences Contextual Cueing of Visual Search." *Front Psychol* no. 7:852. doi: 10.3389/fpsyg.2016.00852.