Mechanistic Versus Functional Understanding

Tania Lombrozo, Department of Psychology, University of California, Berkeley

Daniel Wilkenfeld, Department of History and Philosophy of Science & Center for Philosophy of Science, University of Pittsburgh[1]

Forthcoming in Stephen R. Grimm (Ed.), *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology*. New York, NY: Oxford University Press.

Abstract: Many natural and artificial entities can be predicted and explained both mechanistically, in term of parts and proximate causal processes, as well as functionally, in terms of functions and goals. Do these distinct "stances" or "modes of construal" support fundamentally different kinds of understanding? Based on recent work in epistemology and philosophy of science, as well as empirical evidence from cognitive and developmental psychology, we argue for what we call the "weak differentiation thesis": the claim that mechanistic and functional understanding are distinct in that they involve importantly different objects. We also consider more tentative arguments for the "strong differentiation thesis": the claim that mechanistic and functional understanding involve different epistemic relationships between mind and world.

Keywords: functional explanation, mechanistic explanation, stances, modes of construal, understanding

[1] The authors gratefully acknowledge the John Templeton Foundation for its support through the *Varieties of Understanding* project. We are also grateful to Stephen Grimm for relevant discussions and helpful comments on an earlier draft.

A fawn is born with white spots. An alarm clock beeps unpleasantly. A professor decides to give a pop quiz. How can we understand why these events occurred as they did?

Several psychologists have suggested that events like these support (at least) two forms of understanding – what we will call *mechanistic understanding* and *functional understanding*. Mechanistic understanding relies on an appreciation of parts, processes, and proximate causal mechanisms. The fawn has white spots because of its genes and prenatal environment; the alarm clock beeps when the circuit connecting a power source to a buzzer is completed; the professor decides to give a pop quiz when she sees that her students have not been coming to class prepared. Functional understanding, by contrast, relies on an appreciation for functions, goals, and purpose. The fawn has white spots to hide from predators against sun-flecked ground; the alarm clock beeps to wake its sleeping owner; the professor gives a pop quiz to assess and improve her students' mastery of the course material.

The distinction between mechanistic and functional understanding rests on substantive (if typically implicit) assumptions about what "understanding" amounts to, and about how understanding can be carved up into distinct forms. Our aim in this paper is to evaluate the evidence for mechanistic and functional forms of understanding through the lens of contemporary epistemology and philosophy of science, which offer valuable new tools for thinking about the nature and varieties of understanding. In particular, we evaluate two claims: the *weak differentiation thesis*, according to which mechanistic and functional understanding have importantly different objects, and the *strong differentiation thesis*, according to which mechanistic and functional understanding constitute qualitatively different kinds of understanding.

In Section 1, we briefly introduce a family of related accounts of understanding that have emerged from recent work in philosophy. While we don't commit to a specific member of this family, we take on a shared commitment that guides our subsequent discussion: that understanding is at least partially a matter of representing the right kinds of (explanatory) dependence relationships. In Section 2, we introduce the idea of "stances" or "modes of construal" as cognitive mechanisms that support the construction of mental representations that underwrite this notion of understanding. In Section 3, we review the empirical evidence for the psychological reality of mechanistic and functional modes of construal. In Section 4, we argue for the weak differentiation thesis. Finally, in Section 5, we offer more tentative arguments for the strong differentiation thesis.

1. Understanding as representing (explanatory) dependence

Accounts of understanding within epistemology and philosophy of science differ along a variety of dimensions, including whether understanding is regarded as a type of knowledge (e.g., Grimm 2006), as an ability (e.g., Hills 2016), as possession of a mental model (e.g., Knuuttila & Merz 2009) or as some other form of epistemic relation (Wilkenfeld 2013). However, virtually all extant accounts share one thing in common: because understanding is regarded as a fundamentally cognitive or epistemological relationship, it must be constituted at least in part by how we represent that which is understood.[2] Typically, the contents of these representations are

[2] To our knowledge, the only philosophers who would take issue with this characterization are those who would eschew a representationalist framework on other grounds (e.g., Price 2011). Even on Hills' (2016) ability-centric account, part of *understanding why* is being able to put forward an explanation, which at least *prima facie* seems to require representing the understood as occupying a specific node in an explanatory nexus. Since our focus will be on how the representations that make up understanding are effected, we will concentrate mostly on the representational component—however, everything we say about generating such representations holds true whether they exhaust understanding or only complement some other property of the understander (e.g., an ability).

taken to be of particular kinds of dependence relations, be they causal, explanatory, or counterfactual.

The most straightforward view of the representational content of understanding might be that understanding why P corresponds to having knowledge of the causes of P. This is the picture one gets from the influential work of Woodward (2003), who has a well-worked-out account of the nature of causal information, causal understanding, and causal explanation. Alternatively, one could think that causation is just one species of a more general category of metaphysical dependence relation, knowledge of any of which might constitute understanding (Kim 1994). For instance, non-causal dependence relations could include mereological or conceptual relations.

Another line of thought is that the category of "dependence relations" might be too narrow (or at least not narrow in the correct way) to capture understanding (perhaps because it doesn't capture explanatory connections between necessary truths), and that really the object of the knowledge that constitutes understanding is better thought of as *explanation* generally. The idea is that the representational content of understanding simply *is* knowledge of an explanation. While this view has periodically arisen alongside theories of explanation (most famously Hempel 1965), it has not generally been favored by epistemologists of understanding itself.[3]

The best way to appreciate the source of skepticism regarding purely knowledge-based accounts of understanding is to see what philosophers add to such accounts—frequently a deeper epistemic relation often known as "grasping." One example is what Strevens (2013) refers to as the "simple view," which is the view that understanding is the state one is in precisely when one *grasps* a correct scientific explanation. What the grasping adds to knowledge is that it rules out

[3] There are exceptions—for example, Trout (2007) argues that understanding is redundant with explanation, because the only sort of worthwhile understanding is knowledge of an explanation. (Trout sometimes uses the language of "grasping," but he also suggests (585) that it is really knowledge that he has in mind.) For an extended reply to Trout, see de Regt 2009.

what we might think of as inert knowledge—propositions that could be known without one being able to really see their inferential or practical implications.

While these accounts are distinct, and their advocates propose their own variants, they share a core commitment that will guide our subsequent discussion of the empirical literature. Specifically, these views share the commitment that understanding is at least partially constituted by mental representations that encode the right kinds of dependence relations, where the right kinds of dependence relations are those that are causal, explanatory, or otherwise privileged, perhaps in terms of their functional or inferential roles.

In the section that follows, we'll suggest that "stances" or "modes of construal" support the creation of precisely these kinds of mental representations.

2. Stances or "modes of construal" as paths to understanding

In several influential papers and books, Daniel Dennett introduced the idea of a *stance:* a strategy for interpreting the behavior of an entity (e.g., Dennett 1971, 1989, 2009). Most relevant for our purposes, Dennett differentiated between a physical stance, which involves predicting and explaining the entity through the application of physical laws, and a design stance, which involves predicting and explaining the entity on the basis of its design and proper functioning. For example, someone who predicts what will happen when pressing a button on an alarm clock by considering the underlying electronic components and the physical laws that govern them is applying the physical stance; someone who does so by thinking about how the alarm clock would be designed and assuming that it is functioning properly is applying the design stance.

An idea akin to Dennett's stances was introduced into the psychological literature by Frank Keil, who argued that even young children are equipped with multiple "modes of construal" that "frame" explanations by positing certain kinds of relations, properties, or arguments as central (Keil 2006). Like Dennett, Keil argued that these include a mechanical / physical mode of construal and a teleological / functional mode of construal. Keil argued for these modes of construal on the basis of children's patterns of explanations and predictions across domains (Keil 1994, 1995). In particular, he distinguished between what we will call "mechanistic" explanations (involving parts or proximate causal mechanisms) and "functional" explanations (involving functions, purpose, or goals), where the former reflects the operation of a mechanistic mode of construal, and the latter the operation of a functional mode of construal.

Importantly, modes of construal (or stances; we will use these terms interchangeably) are not themselves domain theories, such as (scientific or intuitive) physics, or (scientific or intuitive) psychology. Nonetheless, these domain theories may be prerequisites to the successful application of a mode of construal: it is these theories that supply the laws required to apply a physical stance, and that constrain inferences about what would constitute good design and proper functioning. Modes of construal, unlike domain theories themselves, provide a template or algorithm of sorts, determining the basis for a prediction or explanation, and accordingly constraining which domain theories will be consulted and how.

If modes of construal are *strategies* for interpreting entities and their behavior, they do not themselves constitute understanding. Nonetheless, we think there are two meaningful ways in which we might say that stances support or reflect understanding, corresponding to the output versus the input to the corresponding mode of construal. First, the representations that result from the application of a stance to a particular entity will include the representational bases for prediction and explanation – typically the identification of causal and explanatory relationships that hold (or are believed to hold) for the case in question. For example, applying the physical stance to an alarm clock will involve representing the components of the alarm clock as instantiating more general causal relationships encoded in an intuitive physical theory. Applying the design stance to a fawn's spots could involve drawing inferences about the function the spots might serve in a particular ecological context. In this way, the application of a stance will include the creation of representations that can constitute at least the representational component of understanding.[4] They can constitute understanding because they satisfy the common requirements for understanding that we identified in Section 1: representing the right kinds of dependence relations.

A second way in which modes of construal might relate to understanding is in the way they pick out aspects of intuitive theories. First, note that an intuitive theory could itself constitutes some form of understanding. On most accounts, intuitive theories are defined in terms of the explanatory generalizations and causal relationships that they represent (e.g., Carey 1985, Gopnik, Maltzoff, & Bryant 1997, Gopnik & Wellman 2012; Murphy & Medin 1985), again providing a good match to the accounts of understanding that we identify in Section 1. On this view, the components of a theory that in a specific context are employed in applying a mechanistic mode of construal might be said to constitute "mechanistic" understanding (in that context), while those employed in applying a functional mode in a specific context might be said to constitute "functional" understanding (in that context).

To illustrate these two ways of relating modes of construal to understanding, consider again our spotted fawn. Someone who has an intuitive theory of biology that includes resources for explaining biological adaptations might possess some "functional understanding" of biological adaptations in general. When applying this to the spotted fawn, she comes to possess

[4] For the remainder of the paper we will talk in terms of constituting understanding, leaving implied the caveat that there might be other necessary conditions on some accounts.

some functional understanding of why fawns have spots. It is this latter form of understanding – the understanding that results from the application of a mode of construal to a particular entity – that the empirical evidence has most closely addressed, and that we turn to in Section 3.

3. The empirical evidence for mechanistic and functional modes of construal

The majority of research on mechanistic and functional modes of construal has focused on mechanistic and functional explanations, with the (often implicit) assumption that the endorsement or generation of each explanation type reflects the operation of its corresponding mode of construal. Accordingly, our review will focus on what we know about these two kinds of explanations.

First, we begin with an important similarity: both mechanistic and functional explanations are understood as causal explanations. For mechanistic explanations this is an uncontroversial claim; they explicitly appeal to proximate causes and causal mechanisms. However, this isn't self-evidently the case for functional explanations – after all, functional explanations explain current properties or events by appeal to potential future consequences, and thus seem to get the causal order wrong. When we explain that the teacher gave a pop quiz "to teach her students a lesson," we seem to be explaining a current action by appeal to an anticipated but unrealized effect of that action. When we explain the fawn's spots by appeal to camouflage, we seem to be explaining a current property by its potential future influence on predators.

Several accounts of functional explanation offer ways to understanding future-looking functional explanations in more standard backwards-looking causal terms (e.g., Allen 2009, Wright 1976). When we explain the teacher's pop quiz by appeal to the goal of teaching the students a lesson, for example, we can take this as a shorthand for the beliefs and desires that were in fact the proximate causes of her action. She *wanted* to teach her students a lesson, and she *believed* that administering a pop quiz would accomplish her goal. These mental states preceded her behavior, and we can understand the functional explanation as a pointer to these antecedent causes. A more complicated but similar move works for adaptationist explanations. When we explain the fawn's spots by appeal to camouflage, this functional explanation is underwritten by a particular set of causal commitments: that spots in fact support camouflage, and that the fact that they do so played a causal role in the maintenance and spread of spots in past fawns that causally led to the existence of the current spotted fawn. Wright (1976) provides a more general formulation of the causal commitments that underwrite functional explanations in terms of what he calls a "consequence etiology."

Critically, empirical evidence supports this analysis of functional explanations as a descriptively adequate account of human cognition. Lombrozo and Carey (2006) presented adult participants with vignettes followed by why-questions and candidate explanations, with the aim of identifying the conditions under which participants would find functional explanations acceptable. They found that Wright's causal commitments were a necessary condition for acceptance, where the relevant causal commitments were both manipulated experimentally and assessed by having participants indicate their agreement with counterfactual claims. Roughly, for some property P to be explained by appeal to some function F, participants had to endorse the claim that had P not resulted in F, the entity with P probably wouldn't have had P.

Additional work supports the idea that functional explanations are tied to particular causal commitments. Kelemen and DiYanni (2005) found that children were more likely to accept a functional explanation for the origins of an entity or event (e.g., "the first ever thunderstorm occurred to give the earth water so everything would grow") if they also believed that the entity or event was created by "someone or something." Adults are also more likely to accept scientifically-unwarranted teleological explanations (e.g., "water condenses to moisten the air") if they endorse some Gaia-like causal force (Kelemen & Rosset 2009; Kelemen, Rottman, & Seston 2013; see also ojalehto, Waxman, & Medin 2013). The evidence thus suggests that while functional explanations may differ from *mechanistic* explanations (as we'll see below), they should not be understood as *non-causal*.

If mechanistic and functional explanations are both causal explanations, we can already see why knowing or grasping them might constitute understanding on the sort of view sketched in Section 1. The next question, then, is how they differ from each other. One differentiating factor has already emerged: whereas mechanistic explanations invoke proximate causal processes directly, functional explanations do so indirectly; they don't wear their causal commitments on their sleeves. But the literature provides two additional bases for differentiation that are worth reviewing in turn: functional explanations are to some extent mechanism-independent, and they have a distinct developmental and cognitive profile.

First, consider the claim that functional explanations are *mechanism-independent* in the sense that they highlight dependence relations that can be multiply realized, and that their explanatory value is enhanced, rather than diminished, by the dissociation from particular mechanisms. The intuition behind these claims is nicely illustrated by William James's description of the relationship between Romeo and Juliet (an intentional system) versus iron filings and a magnet (a physical system):

> "Romeo wants Juliet as the filings want the magnet. And if no obstacles intervene, he moves toward her by as straight a line as they. But Romeo and Juliet, if a wall be built between them, do not remain idiotically pressing their faces against its opposite sides, as in fact the iron filings do, pursuing the magnet. Romeo soon finds a circuitous way, by scaling the wall or otherwise, of touching Juliet's lips directly. With the filings the path is fixed; whether it reaches the end depends on accidents. With the lover it is the end which is fixed, the path may be modified indefinitely" (James 1890, p. 20).

Romeo, unlike the iron filings, will find an alternative way to reach Juliet. He'll climb the wall; he'll dig a tunnel. The relationship that's stable is that between Romeo's goal of reaching Juliet and his eventual arrival at her side; the means by which he accomplishes this might be variable and highly contingent on idiosyncratic features of the way things happened to unfold. It's this sense in which reasoning about Romeo and Juliet in terms of functional relationships is mechanism (or means) independent. Correspondingly, we can explain Romeo's actions with a functional explanation ("he went that way *to reach Juliet*"), and this might strike us as more appropriate than a mechanistic explanation ("he went that way because he moved his muscles in such and such a way…") precisely because it identifies the dependence relation that's robust across irrelevant perturbations (see also Murray & Lombrozo 2017).

Consider an example from Daniel Dennett, motivating the design stance:

> "Suppose I categorize a novel object as an alarm clock: I can quickly reason that if I depress a few buttons just so, then some hours later the alarm clock will make a loud noise. I don't need to work out the specific physical laws that explain this marvelous

regularity; I simply assume that it has a particular design—the design we call an alarm clock—and that it will function properly, as designed." (Dennett 1991)

In a case like this, the relationship between the buttons and the noises can be multiply realized; it's the function or design of the clock that constrains their relationships. We don't need to reason on the basis of physical laws and causal mechanisms because the explanatory and predictive relationships that we care about are mechanism independent. "The essential feature of the design stance," Dennett writes, "is that we make predictions solely from knowledge or assumptions about the system's functional design, irrespective of the physical constitution or condition of the innards of the particular object" (Dennett 1971). This is part of what makes the design stance so powerful: we can achieve some predictive and explanatory competence without detailed knowledge of general mechanisms or detailed knowledge of how particular causal processes unfolded in the past.

Psychological evidence supports the idea that while functional explanations are understood as causal explanations, they are (at least somewhat) mechanism independent. One source of evidence comes from studies that have examined people's patterns of generalization, where they could generalize on the basis of proximate mechanisms or on the basis of functions and design (Ahn, 1998; Lombrozo, 2009; Lombrozo & Gwynne, 2014; see also Lombrozo & Rehder, 2012). When participants were given or generated functional explanations, they were significantly less likely to generalize on the basis of proximate mechanisms (relative to functions).

To illustrate, consider a study from Lombrozo and Gwynne (2014). In this study, participants learned about animals and artifacts, where each had a target property that could be explained either mechanistically or functionally. For example, some participants read about a plant called a narp with a speckled pattern. They learned that "biologists have discovered that in narps, the speckled pattern is caused by the XP2 gene." This supported the mechanistic explanation that narps have a speckled pattern because of the gene. They also learned that "having a speckled pattern attracts butterflies, which play a role in pollination." This supported the functional explanation that narps have a speckled pattern to attract butterflies for pollination. Participants were then asked to explain, in a sentence, why narps have a speckled pattern. This prompt was deliberately ambiguous: it could be answered by providing a mechanistic explanation, a functional explanation, or both.

After responding to the ambiguous prompt, participants learned about novel items that shared either the proximate cause (e.g., another plant with the XP2 gene) or the function (e.g., another plant that attracts butterflies for pollination), and they were asked whether they would generalize properties from the initial item (the narp) to these new cases. For example, if participants were told that the speckled pattern on narps is high in contrast, would they be inclined to think that the speckled pattern of the other plant with the XP2 gene, or the other plant that attracts butterflies, was also high in contrast? A key finding was that for biological organisms, those participants who provided a functional explanation in response to the ambiguous prompt were less likely than those who did not do so to generalize on the basis of underlying causal mechanisms. Instead, for all types of items, participants who provided a functional explanation were more likely than those who did not do so to generalize on the basis of shared functions.

A second source of evidence for the idea that some level of mechanism independence can be induced by a functional mode of construal comes from people's causal ascriptions. Lombrozo

(2010) presented participants with vignettes in which three causal factors interacted to bring about an effect. For example, in one vignette, participants read about a type of shrimp that eats three food sources, call them A, B, and C. They further learned that eating these three food sources results in the shrimp reflecting high frequencies of UV light. The causal relationships between these factors were designed to create a situation involving "double prevention": A could cause the shrimp to reflect high frequencies of UV light on its own, if not prevented by B. But C prevented B from preventing A, thereby resulting in the effect. This causal structure was used to isolate a notion of causation based on counterfactual dependence from one based on a physical mechanism involving what philosophers often call production or transmission (e.g., Hall 2004). Specifically, while C "caused" the effect in the sense that the effect would not have occurred in its absence, C did not produce the effect through some spatiotemporally continuous mechanism or direct transmission of force.

The key experimental manipulation was whether participants were given additional information that would allow them to construe the relationship between C and the effect functionally. Half the participants were told that the effect (reflecting high frequencies of UV light) serves a biological function (temperature regulation), and that the shrimp evolved to eat A and C for this reason. The key finding was that participants were significantly more inclined to consider C a cause of the effect when this functional relationship held, such that the difference in ratings between A (the productive cause) and C (the dependence cause) was decreased. This suggests that when construing a relationship functionally, participants' judgments of whether some factor caused an effect were less sensitive to the nature of the mechanism mediating the counterfactual dependence between the effect and the candidate cause.

A third source of evidence that functional thinking induces some insensitivity to mechanistic information comes from the Illusion of Explanatory Depth, or IOED (Rozenblit & Keil 2002). The basic finding is that people tend to overestimate their mechanistic understanding of how devices such as helicopters or flush toilets work. One suggestion is that people mistake a *functional* understanding for how something was designed or operates for a mechanistic understanding of the actual causal processes involved. Consistent with this idea, Alter, Oppenheimer, and Zemla (2010) found that when participants adopted a more abstract mode of construal, which is itself associated with reasoning in terms of functions, they experienced a larger IOED. When reasoning functionally, it seems, they had less metacognitive access to their deficient mechanistic understanding.

A fourth source of evidence for a relationship between a functional mode of construal and mechanism independence comes from looking-time studies with infants. Woodward and her colleagues have shown that when infants construe an agent's action as a goal-directed reach, they are more likely to expect that the agent's next action will preserve the same goal, even if it involves a departure in means, such as reaching left versus right (e.g., Cannon & Woodward 2012, Woodward 1998). Research by Gergely and colleagues illustrates that infants can also use variation in means as a basis for inferring that an agent is rational in its pursuit of goals: when an agent's goal is preserved despite variation in means, 12-month-olds develop expectations that the agent will seek the goal, and will do so in the most rational (i.e., spatially efficient) way possible (e.g., Gergely, Nadasdy, Csibra, & Biro 1995).

To sum up, these studies on generalization, causal ascription, metacognition, and infants' perception of goal-directed action support the idea that adopting a functional construal differs from a mechanistic construal in that the former allows for a more mechanism-independent form of reasoning. More recent work by Liquin and Lombrozo (in prep) sheds further light on why this might be. They find that when evaluating a functional explanation, judgments are largely (though not exclusively) driven by an assessment of how well a proposed feature (such as "reflecting high frequencies of UV light") "fits" a given function (such as "thermal regulation"). This evaluation of structure-function fit may involve some mechanistic reasoning, but it crucially *does not* depend upon a detailed analysis of the feature's etiology. Indeed, Liquin and Lombrozo find that when an explanation contains functional information, participants become less sensitive to etiological detail.

A second factor that differentiates mechanistic and functional construals may or may not be related: there's evidence that functional explanations may be psychologically privileged in the sense that they are often favored and seem to be less cognitively demanding. In particular, there's evidence that children use them "promiscuously" (Kelemen 1999), and that adults will accept scientifically-unwarranted functional explanations when cognitively impaired (Lombrozo, Kelemen, & Zaitchik 2007), when responding under speeded conditions (Kelemen & Rosset 2009; Kelemen, Rottman, & Seston 2013), or when they engage in less reflective thought (Steiner, Zemla, & Sloman 2016). Kelemen and colleagues have argued that a teleological mode may be a "cognitive default" that emerges early in development and remains throughout the lifespan, re-emerging when alternative cognitive resources are taxed (Kelemen, Rottman, and Seston 2013; see also Shtulman & Lombrozo 2016). Liquin and Lombrozo (in prep) argue that this is because structure-function fit serves as an intuitive but defeasible heuristic cue to the acceptability and quality of a functional explanation (see also Lombrozo, Kelemen, & Zaitchik, 2007).

The evidence reviewed so far suggests that mechanistic explanations differ from functional explanations in their causal commitments, that mechanistic and functional explanations support different patterns of generalization and causal ascription, and that functional explanations may be cognitively privileged in some sense. These forms of differentiation suggest that mechanistic and functional modes of construal could be tuned to different inferential functions. Rather than adopting a mechanistic mode of construal for some kinds of objects and a functional mode of construal for others, people could flexibly adopt one mode or the other depending on the entity in question and their inferential aims. Indeed, there's some evidence that this is the case. Not only do people spontaneously offer one kind of explanation or another in response to various features of the entity in question (e.g., Lombrozo & Carey 2006; Sanchez et al. 2016), they also adapt their evaluations to their inferential goals: they rate mechanistic explanations more highly when they anticipate making inferences on the basis of proximate mechanisms, and they rate functional explanations more highly when they anticipate making inferences on the basis of function (Vasilyeva, Wilkenfeld, & Lombrozo 2017).

In sum, there is good evidence for the claim that mechanistic and functional explanations are psychologically distinct. They not only differ in their causal commitments, but also in the extent to which they demand and depend upon an articulation or specification of mechanisms or particular causal processes. They also differ in the dependence relations that they privilege for the purpose of generalization. Perhaps for these reasons, functional explanations seem to have a special role in development and may be less cognitively demanding.

On the view of understanding articulated in Section I, representations of mechanistic and functional explanations are good candidates for understanding: they encode causal and explanatory relationships that plausibly support an understanding for why some entity has a particular property or exhibited a particular behavior. But do they merely support understanding of different things? Or do they constitute different kinds of understanding? We turn to the weak and strong differentiation theses in Sections 4 and 5, respectively.

4. The case for *weak differentiation*

In Section 1, we suggested that understanding involves some representation of dependence or explanatory relations. In Sections 2 and 3, we suggested that mechanistic and functional modes of construal support mechanistic and functional understanding, respectively. In this section, we consider whether this evidence supports the weak differentiation thesis, namely that mechanistic and functional understanding are different insofar as they involve different objects (whether or not they also involve different epistemic relations to those objects). We suggest that mechanistic and functional understanding indeed involve different content, support different functions, and have a distinctive phenomenology. However, we will also argue that each of these claims is insufficient to support the strong differentiation thesis that these constitute different *kinds* of understanding.

The claim that mechanistic and functional understanding involve different content follows straightforwardly from the data presented in Section 3. We've seen that they involve mechanistic versus functional explanations, privilege production versus dependence notions of causation, and privilege different dependence relations as a basis for inference. Yet there are good reasons to doubt that understanding should simply inherit criteria for individuation from explanation, causation, or some inferential role. If understanding is a relation between mind and world, it might be the same relation even when the world provides starkly different relata. As an analogy, knowledge of mechanisms and knowledge of functions take very different objects, but we would not for that reason usually be inclined to say that they manifest more than one knowledge relation. There might be some objects so diverse that we have reason to posit multiple knowledge relations, but mechanisms and functions can still be known in approximately the same way.

Mechanistic and functional understanding also differ with respect to their core functions. While both support prediction and explanation, mechanistic understanding is particularly useful for prediction and explanation in some domains, while functional understanding is more useful in others. Moreover, as shown in Vasilyeva, Wilkenfeld, and Lombrozo (2017), people privilege the explanations that support their current inferential goals. But once again, it's not clear that supporting different *kinds* of inferences underwrites the stronger claim that an understanding of whatever-supports-mechanistic-inferences and an understanding of whatever-supports-functional-inferences are different *kinds* of understanding. Knowledge of statistics supports inductive inferences, whereas knowledge of geometry supports deductive inferences, but we would not on that basis typically be inclined to consider them different *kinds* of knowledge.

Finally, consider the claim that mechanistic and functional understanding are distinct with regard to their phenomenology. This claim goes admittedly beyond the data, but it's only a modest step from the claim that functional explanations are a cognitive default of some kind (a claim that may or may not be right) to the claim that they are satisfying in a more basic or intuitive way. However, there is reason to doubt that when two tokens of understanding feel different to their respective understanders, we have good grounds for saying that they belong to two different *kinds* of understanding. Knowledge that one is in danger might feel quite different from knowledge that one is safe; it doesn't follow that the knowledge relation is different in kind.

If differences in content, function, and phenomenology are insufficient to support the claim that mechanistic and functional understanding are different kinds of understanding, it might be tempting to reject their uniqueness entirely, and to instead consider the possibility that mechanistic and functional understanding are but two among a very large number of possible targets for understanding. On this licentious view, any strategy for privileging a subset of the enormously complex (explanatory) dependence relations in the world offers a "mode of construal" and thus a possible target for understanding. Moreover, mechanistic and functional understanding have no special status with respect to these alternatives, and all of these alternatives support understanding in just the same way: by supporting representations of the dependence relations that constitute understanding.

We think this possibility misses something important. It's not a coincidence that mechanistic and functional stances or construals arise again and again in philosophy and in psychology, across disciplines and over time. These two construals – unlike an arbitrary subset of dependence relations – seem to capture something important about the structure of the world and our goals within it. Proximate causes and goals, under the right circumstances, identify dependence relations that are particularly stable, or insensitive to perturbations in background conditions (Lombrozo 2010, see also Blanchard, Vasilyeva, & Lombrozo forthcoming, Woodward 2006). Given our goals, they might be particularly useful bases for prediction and intervention. For these reasons, it seems appropriate to recognize mechanistic and functional understanding as understanding of special kinds of targets, even if the understanding itself is not different in kind.[5] It's for this reason that we favor some form of differentiation between mechanistic and functional understanding, even if it's only a weak form.

In sum, we think there is good evidence for the weak differentiation thesis: mechanistic and functional understanding have objects that are both important and importantly different from each other. At the same time, we don't think that the evidence just reviewed supports the stronger claim that mechanistic and functional understanding involve qualitatively different kinds of understanding. In Section 5, we evaluate this stronger claim.

5. The case for *strong differentiation*

In this section, we consider two tentative arguments for the strong differentiation thesis: the claim that mechanistic and functional understanding reflect different epistemic relationships to the world. The two arguments that we consider are that mechanistic and functional understanding differ in their normative entailments and that they differ in their modal implications. These arguments are tentative in part because they stem from intuitive considerations rather than fully developed theoretical arguments, and in part because they have empirical commitments that have yet to be tested. Nonetheless, we think these possibilities merit further study, and so we sketch them here.

First, mechanistic and functional understanding seem to differ with regard to normative considerations. When we learn that an alarm clock has the function of waking its owner, we're in a position to evaluate whether it has done so well. When we learn that a fawn has spots for camouflage, we can evaluate how it might better accomplish this goal. A mere causal process, on

[5] We don't mean to suggest here that functional and mechanistic understanding are the only targets of understanding that might have this special status. There's already good evidence for the psychological reality of something like intentional (e.g., Bertram & Hodges 2005) and formal (e.g., Prasada 2016) modes of construal, which could either be special targets of understanding or perhaps constitute different kinds of understanding.

the other hand, does not support normative evaluations, at least not on its own. Causal processes—absent any reference to goals—are simply facts about the world, with no standard available against which to compare them. In David Hume's *Treatise of Human Nature* (Book III), he famously puts forward the claim that you cannot derive claims about what *ought to be* from any number of premises about what *is*. We cannot even rightfully say that a causal process is effective without making assumptions regarding what counts as the relevant effect.[6]

Functional explanations also differ from mechanistic explanations in that the normative evaluations supported by the former involve an implicit perspective or point of view (see also ojalehto, Waxman, & Medin 2013, for the idea that functional explanations are importantly perspectival). When we evaluate the alarm clock, we do so from the perspective of the designer or the user. When we evaluate the fawn's spots, we do so from the perspective of the fawn. (Presumably, effective camouflage is *not* a desirable characteristic for a fawn from the perspective of a mountain lion.) Again, causal processes are simply features of the world – they do not, on their own, offer a perspective or point of view for further evaluation.

If functional understanding involves normativity and an implicit perspective, but mechanistic understanding does not, we have the first hints that the relationship between the mind and the world may differ across these cases of understanding. Whereas mechanistic understanding involves a mind-to-world fit (like a *belief*), functional understanding additionally has elements of a world-to-mind fit (like a *desire*). Functional understanding involves a perspective from which one can appreciate how the world would be more desirable (from that perspective).

---

[6] We might be able to evaluate which of two processes is more "efficient" in the purely technical sense of "using less energy to accomplish the same task." However, to see that this still has no normative implications, we need only imagine a goal for which the less efficient process is better—for example, in using up a department's budget so that it is not slashed for the following fiscal year.

This aspect of functional understanding parallels an aspect of *first-personal* understanding as developed by Grimm (2016). Grimm's aim is to articulate how understanding the action of another person differs from understanding the structure of the natural world. Grimm argues that in the former case, it's not enough to identify some of the (causal) structure upon which an action depends; it's also important to understand why the action was desirable, or the choice choiceworthy. He writes that "this sort of seeing [of a goal as desirable or choiceworthy] plausibly requires a different cognitive attitude—and hence, apparently, a different cognitive method—that we need to draw upon when we try to understand other human beings." This is part of what motivates Grimm's conclusion that understanding another person is different from third-personal understanding, which depends only on grasping the right kind of dependence structure (e.g., between a desire and some action), without the further step of understanding not only why it is *desired*, but why it is *desirable*. Similarly, functional understanding plausibly involves different "cognitive methods" from mechanistic understanding, though in the case of functional understanding, it suffices to understand what is desirable from the implicit perspective of a functional explanation, without it being desirable in Grimm's 1st-personal sense.

The second way in which mechanistic and functional understanding could differ is with respect to the specificity of their commitments regarding the causal structure of the world. As we argued in Section 3, functional understanding is – in an important sense – mechanism independent. When we obtain functional understanding, our causal commitments radically underdetermine the actual causal process by which some property or event came to be. We can functionally explain why the alarm clock beeped by appealing to its design – and be satisfied with our explanation – even if we remain forever ignorant of whether its inner parts function electronically or pneumatically. For functional explanations, a "how possibly" story goes a long way; it might be enough to know that *some* process with a consequence etiology (Wright 1976) was at work, without knowing more about what it was or how it manifested. The same can't be said for mechanistic understanding. We might be satisfied by a vague appeal to the alarm clock's internal electronic processes, but a mechanistic explanation seems to demand a "how actually" story. If this is right, then mechanistic and functional understanding are qualitatively distinct in the sense that they are differentially demanding with respect to what the actual causal structure of the world must be like.

In sum, we've sketched two arguments for the strong differentiation thesis. We've suggested that mechanistic and functional understanding involve different epistemic relationships in that the latter has normative and perspectival elements that introduce a world-to-mind fit. We've also argued that because functional understanding is mechanism-independent, it makes weaker demands on the causal structure of the world – *possibility* is enough for understanding. These claims go beyond the weak differentiation thesis because they posit that mechanistic and functional understanding differ not only in terms of their objects, but in the mind-world relation that they require.

6. **Conclusion**

Our aim in this paper has been to review empirical evidence for mechanistic and functional modes of construal, and to relate this evidence to accounts of understanding. First, we argued that these modes of construal support understanding because they play a role in generating the kinds of representations that (at least partially) constitute understanding. Next, we argued that mechanistic and functional understanding are distinct in two ways: they involve importantly different objects, and (more tentatively) they involve different epistemic relationships. These claims have implications for how to think about understanding in epistemology and philosophy of science. They also invite us to ask a host of empirical questions about the psychological capacities that underwrite these forms of understanding, and about their implications for our interactions with the world.

References

Ahn, Woo-kyoung. 1998. "Why are Different Features Central for Natural Kinds and Artifacts?: The Role of Causal Status in Determining Feature Centrality." *Cognition* 69(2): 135-178.

Allen, Colin, "Teleological Notions in Biology", *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), Ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/win2009/entries/teleology-biology/>.

Alter, Adam L., Oppenheimer, Daniel M., and Zemla, Jeffrey C. 2010. "Missing the Trees for the Forest: A Construal Level Account of the Illusion of Explanatory Depth." *Journal of Personality and Social Psychology 99*(3): 436-451.

Cannon, Erin N., and Woodward, Amanda L. 2012. "Infants Generate Goal-Based Action Predictions." *Developmental Science* 15(2): 292-298.

Carey, Susan. 1985. *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.

Dennett, Daniel C. 1971. "Intentional Systems." *The Journal of Philosophy* 68(4): 87-106.

-----. 1989. *The Intentional Stance*. Cambridge, MA: MIT press.

-----. 2009. "Intentional Systems Theory." In *The Oxford Handbook of Philosophy of Mind*. Eds. Ansgar Beckermann, Brian P. McLaughlin, and Sven Walter. New York: Oxford University Press.

De Regt, Henk. W. 2009. "The Epistemic Value of Understanding." *Philosophy of Science 76*(5): 585–597.

Gergely, György, Nádasdy, Zoltán, Csibra, Gergely, and Bíró, Szilvia. 1995. "Taking the Intentional Stance at 12 Months of Age." *Cognition* 56(2): 165-193.

Gopnik, Alison, Meltzoff, Andrew N., and Bryant, Peter. 1997. *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Gopnik, Alison, and Wellman, Henry. M. 2012. “Reconstructing Constructivism: Causal Models, Bayesian Learning Mechanisms, and the Theory Theory.” *Psychological Bulletin 138*(6): 1085-1108.

Grimm, Stephen. R. 2006. “Is Understanding a Species of Knowledge?” *The British Journal for the Philosophy of Science 57*(3): 515–535.

-----. 2016. "How Understanding People Differs from Understanding the Natural World." *Philosophical Issues* 26(1): 209-225.

Hall, Ned. 2004. "Two Concepts of Causation." In *Causation and Counterfactuals*: Eds. John Collins, Ned Hall, and Laurie Paul. Cambridge, MA: MIT Press.

Hempel, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press.

Hills, Alison. 2016. Understanding Why. *Noûs*, *50*(4): 661–688.

Hume, David. [1738/1740] 2003. *A Treatise of Human Nature*. Courier Corporation.

James, W. 1890. The principles of psychology (Vol. 1). *New York: Holt*, *474*.

Keil, Frank C. 1994. "The Birth and Nurturance of Concepts by Domains: The Origins of Concepts of Living Things." In *Mapping the Mind: Domain Specificity in Cognition and Culture*: Eds. Lawrence A. Hirschfeld and Susan A. Gelman. Cambridge: Cambridge University Press.

-----. 1996. "The Growth of Causal Understandings of Natural Kinds." In *Causal Cognition: A Multidisciplinary Debate*: Eds. Dan Sperber, David Premack, and Ann James Premack. New York: Oxford University Press.

-----. 2006. "Explanation and Understanding." *Annual Review of Psychology.* 57: 227-254.

Kelemen, Deborah. 1999. "Function, Goals and Intention: Children's Teleological Reasoning About Objects." *Trends in Cognitive Sciences* 3(12): 461-468.

Kelemen, D. and DiYanni, C., 2005. Intuitions about origins: Purpose and intelligent design in children's reasoning about nature. *Journal of Cognition and Development*, *6*(1), pp.3-31.

Kelemen, Deborah, and Rosset, Evelyn. 2009. "The Human Function Compunction: Teleological Explanation in Adults." *Cognition* 111(1): 138-143.

Kelemen, Deborah, Rottman, Joshua, & Seston, Rebecca. 2013. "Professional Physical Scientists Display Tenacious Teleological Tendencies: Purpose-Based Reasoning as a Cognitive Default." *Journal of Experimental Psychology: General* 142(4): 1074-1083.

Kim, Jaegwon. 1994. "Explanatory Knowledge and Metaphysical Dependence." *Philosophical Issues 5*: 51–69.

Knuuttila, Tarja, and Martina Merz. 2009. "An Objectual Approach to Scientific Understanding: The Case of Models." In *Scientific Understanding: Philosophical Perspectives*, edited by Henk De Regt, Sabina Leonelli, and Kai Eigner, 146–168. University of Pittsburgh Press.

Lombrozo, Tania. 2009. "Explanation and Categorization: How "Why?" Informs "What?"." *Cognition* 110(2): 248-253.

Lombrozo, Tania, and Carey, Susan. 2006. "Functional Explanation and the Function of Explanation." *Cognition* 99(2): 167-204.

Lombrozo, Tania, and Gwynne, Nicholas Z. 2014. "Explanation and Inference: Mechanistic and Functional Explanations Guide Property Generalization." *Frontiers in Human Neuroscience* 8: 700.

Lombrozo, Tania, Kelemen, Deborah, and Zaitchik, Deborah. 2007. Inferring Design: Evidence of a Preference for Teleological Explanations in Patients with Alzheimer's Disease. *Psychological Science 18*(11): 999-1006.

Lombrozo, Tania, and Rehder, Bob. 2012. Functions in Biological Kind Classification. *Cognitive Psychology 65*(4): 457-485.

Liquin, Emily, and Lombrozo, Tania. 2017. Structure-function fit in the evaluation of teleological explanations. *Manuscript in preparation.*

Malle, Bertram F., and Hodges, Sara, D. 2005. *Other Minds.* New York: Guilford Press.

Murphy, Gregory. L., and Medin, Douglas. L. 1985. The Role of Theories in Conceptual Coherence. *Psychological Review 92*(3): 289.

Murray, D. and Lombrozo, T. 2017. Effects of manipulation on attributions of causation, free will, and moral responsibility. *Cognitive science*, *41*(2), pp.447-481.

ojalehto, bethany, Waxman, Sandra R., and Douglas L. Medin. 2013. "Teleological Reasoning About Nature: Intentional Design or Relational Perspectives?" *Trends in Cognitive Sciences* 17(4): 166-171.

Prasada, Sandeep. 2016. “Mechanisms for Thinking About Kinds, Instances of Kinds, and Kinds of Kinds.” In *Core Knowledge and Conceptual Change:* Eds. David Barner and Andrew S. Baron. New York: Oxford University Press.

Price, Huw. 2011. *Naturalism Without Mirrors*. New York: Oxford University Press.

Rozenblit, Leonid, and Keil, Frank. 2002. “The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth.” *Cognitive Science 26*(5): 521-562.

Sánchez Tapia, Ingrid, Gelman, Susan A., Hollander, Michelle A., Manczak, Erika M., Mannheim, Bruce, and Escalante, Carmen. 2016. "Development of Teleological Explanations in Peruvian Quechua-Speaking and US English-Speaking Preschoolers and Adults." *Child Development* 87(3): 747-758.

Shtulman, Andrew, and Lombrozo, Tania. 2016. “Bundles of Contradiction: A Coexistence View of Conceptual Change.” In *Core Knowledge and Conceptual Change:* Eds. David Barner, and Andrew S. Baron. New York: Oxford University Press.

Steiner, Samantha M., Zemla, Jeffrey C., and Sloman, Steven. 2016. "MP10: Analytical Style Predicts Religious and Teleological Belief." *Journal of Investigative Medicine* 64(3): 808-809.

Strevens, Michael. 2013. “No Understanding Without Explanation.” *Studies in History and Philosophy of Science Part A* 44(3): 510–515.

Trout, J. D. 2007. “The Psychology of Scientific Explanation.” *Philosophy Compass* 2(3): 564–591.

Vasilyeva, Nadya, Wilkenfeld, Daniel, and Lombrozo, Tania. 2017. "Contextual Utility Affects the Perceived Quality of Explanations." *Psychonomic Bulletin & Review*: 1-15.

Woodward, Amanda L. 1998. "Infants Selectively Encode the Goal Object of an Actor's Reach." *Cognition* 69(1): 1-34.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

-----. 2006. "Sensitive and Insensitive Causation." *The Philosophical Review* 115(1): 1-50.

Wright, Larry. 1976. *Teleological Explanations: An Etiological Analysis of Goals and Functions*. Berkeley, CA: University of California Press.