Psychological Inquiry, 22: 108–114, 2011

ISSN: 1047-840X print / 1532-7965 online
DOI: 10.1080/1047840X.2011.547387

Psychology Press
Taylor & Francis Group

HIT on the Psychometric Approach

Daniel Burnston, Benjamin Sheredos, and William Bechtel
Department of Philosophy, University of California, San Diego, La Jolla, California

Introduction

Traditionally, identity and supervenience have been proposed in philosophy of mind as metaphysical accounts of how mental activities (fully understood, as they might be at the end of science) relate to brain processes. Kievet et al. (this issue) suggest that to be relevant to cognitive neuroscience, these philosophical positions must make empirically testable claims and be evaluated accordingly—they cannot sit on the sidelines, awaiting the hypothetical completion of cognitive neuroscience. We agree with the authors on the importance of rendering these positions relevant to ongoing science. We disagree, however, with their proposal that a metaphysical relationship (identity or supervenience) should "serve as a means to conceptually organize and guide the analysis of neurological and behavioral data" (p. 69). Instead, we advance a different view of the goals of cognitive neuroscience and of the proper means of relating metaphysics and explanation.

Our central objection to the psychometric approach deployed by Kievet et al. is that the formal models only account for *correlations between variables* (measurements) and do not aid in *explaining phenomena*. Cognitive neuroscience is concerned with the latter. We develop this point in the next section, in which we present what we find to be problematic in their proposed models. In the Identity Claims and Mechanistic Explanations in Cognitive Neuroscience section, we advance an account of what is required to explain phenomena: (a) providing an adequate description of a phenomenon, and (b) characterizing the mechanism responsible for it. In doing so we characterize a version of the identity theory—heuristic identity theory (HIT), which figures centrally in developing such explanations—and illustrate its role in what we take to be a prototypical example of research in cognitive neuroscience. Finally, in the Levels, Mechanisms and Identity Claims section, we turn to how levels and interlevel relations should be construed in a metaphysical account that fits the mission of cognitive neuroscience.

Explanation of Correlations Versus Explanation of Phenomena

To situate our critique, we draw upon Bogen and Woodward's (1988) distinction between data, which are measured, and phenomena, which are repeatable processes in the world that are to be explained. Measurements can provide an epistemic inroad to phenomena but should not be confused for phenomena themselves. Reading ability is a cognitive phenomenon; response time on the Stroop task is often used as a measure of that phenomenon. Attention is a cognitive phenomenon; success at discriminating targets from distracters is one measure of it. Memory is a cognitive phenomenon; accurate recall is often used to measure it. If one examines any research area in cognitive neuroscience, it is clear that the researchers are trying to explain phenomena such as memory or attention (albeit typically more finely delineated). Any importation of metaphysical theories should be tightly bound up with such explanation of phenomena, *not* solely with the correlation of variables that results from measurements. In contrast, Kievit et al.'s project seeks to account for correlations between types of measurements via purely formal/mathematical descriptions of their relationship. They then propose to test whether identity or supervenience is better supported by specific models of the correlations between measured variables. This is not in tune with the explanatory endeavors of cognitive neuroscience.

We consider first their treatment of identity theory. The authors propose to formalize identity theory with "reflective models," in which the states of multiple "indicators" are represented in a structural equation as "caused by" the state of a common latent variable (see Figure 3 and p. 73 of their article). Each indicator is provided a specific value via some measurement. In the authors' case studies, the values of psychological or "P-indicators" are specified by measurements such as scored performance on a task. The neurological or "N-indicator(s)" are measurements such as brain mass volume or gray matter density. An identity claim is viewed as justified when a reflective model has a high degree of fit to both sets of data; with a high degree of fit, one may infer that one is "*measuring the same thing*" with both P- and N-indicators (p. 73). It is the structural equation that shows that the causes of the two variables are *identical*: the cause of each is the value of the latent variable. When the reflective model has a good fit to the data, it is hypothesized that "the indicators measure the same thing," where this implies (perhaps incorrectly) that "the latent variable or attributes exists independently of the model specification" (p. 72).

What Kievet et al. offer is a metaphysical account of *how* two sets of measurements can be correlated. But, as previously noted, cognitive neuroscience is not in the business of accounting for correlated data sets; rather, its goal is to explain cognitive phenomena themselves. The latent variable identified in a formative model does not play a role in this type of explanation. In part, this is because it *has no positive description*. The authors' only characterization of the latent variable is as "cause of the indicators." Yet even this is merely an interpretation: All that is inherent in the model is a formal description of the latent variable's *relation to the indicators* (Borsboom, Mellenbergh, & van Heerden, 2003). The important point to note here is that the latent variable is defined *only* with respect to this relation, and therefore there is no characterization of any cognitive or brain process at work *separate from* the indicators. Without such a description, the target of the explanation cannot in fact be the phenomenon of interest but only the data itself. Given that there is a real distinction between measurements and phenomena, the authors' models, and the latent variables that are supposed to pull metaphysical weight in them, are only accounting for the measurement side of the distinction.[1] If we are right in characterizing cognitive neuroscience, and the metaphysical commitments it makes, in terms of explaining phenomena, then the authors' models are not being invoked to construct a genuine metaphysics of cognitive neuroscience at all.[2] Even given a successful reflective model, if the explanatory questions that cognitive neuroscientists *actually ask* are broached—that is, *what is the phenomenon at issue?* and *how does it come about?*—the reflective model has no answer, other than to say that the "indicators [of the phenomenon] can be said to be on equal empirical footing in that they are both assumed to be imperfect reflections of the true state" of the underlying cause (p. 73). We are left waiting for a metaphysical account of cognitive phenomena.

Consider now the authors' treatment of supervenience theory. The model the authors provide for importing the metaphysics of supervenience is a "formative model" in which the latent variable is represented in the structural equation as a *function of* the values of multiple indicators.[3] In the context at hand, the possible values of the latent variable are functions of the values of the N-indicators. This provides a strong "bottom-up" constraint: At a time of measurement, the actual value of the latent variable is "*determined by* a weighted summation" of the values assigned to the N-indicators (p. 75; see also Figure 4). The values of the P-indicators, in turn, are represented as determined by the value of the latent variable. This provides a weaker, "top-down" constraint on what values the latent variable can take (hence, which values the N-indicators can take) if the model is to account for the data. The authors take this to provide a formal rendering of the *dependence* of supervenient psychological properties on subvenient neural properties: If two individuals have the same values for the same N-indicators, then (according to the model) they will have the same values for the same P-indicators. The model is also taken to provide a formal rendering of *multiple realizability*: Two individuals may have the same values for the P-indicators but may nonetheless have different values for the N-indicators.

The same points we have raised against the reflective model serve to illustrate our core complaint against the formative model.[4] Because the latent variable in

[1]There are occasions when scientists do explain data—for example, when they suspect that the data are an artifact of experimental procedures (Bogen & Woodward, 1988). However, this is a very different type of explanation than is the primary focus of cognitive neuroscience. We therefore adopt the "accounting for" language when the focus is on measurements to avoid conflation of very different contexts that might both be broadly "explanatory."

[2]By interpreting the latent variable *causally,* the authors are making *a* metaphysical claim, but it is not the sort of ontological commitment that can help explain cognitive phenomena. In effect, there is not *enough* metaphysics involved—explanation involves making a commitment to *particular* phenomena with particular inherent properties, not simply describing a small set of external relations that may or may not track anything going on in the world.

[3]The authors employ a subtype of formative model, called a "MIMIC" model, which stands for Multiple Indicators, Multiple Causes. Notably, however, the relationship involved in supervenience is one of *realization*, not of causation. The authors are thus providing a novel interpretation of the MIMIC model, which renders its traditional title misleading in this context.

[4]We have two concerns that are specific to the treatment of latent variables in formative models. The first involves the metaphysical status of the latent variables. There is an ongoing debate regarding the ontological status of latent variables, in which one of the authors of the target article is a participant (Borsboom, 2008; Borsboom et al., 2003). The status of latent variables in formative models is particularly vexed, because (as admitted in the target article) in formative models, "the latent attribute is defined by the choice of predictors" and "a change of predictors implies a change in the nature of the attribute" (p. 72). As a result, scholars disagree over whether to take a broadly *realist* or *antirealist* interpretation of latent variables in formative models. This is a topic one would hope to see further addressed by the authors, because it has serious consequences for a metaphysical interpretation of the model. If it is best to adopt an antirealist approach to the latent variable in formative models, then clearly the metaphysical goals of the authors will be unsatisfied. Second, it is unclear whether the formative model actually captures multiple realizability. The authors state that "two people can have different indicator values but the same position at the latent attribute level. Therefore the position on the theoretical attribute is *multiply realizable*" (p. 74). This statement is doubly perplexing. First, if the latent variable *just is* some function of the neural values, it would seem to be itself neural in nature, but traditionally, supervenience theory has concerned the multiple neural realizability of *psychological* features. The authors speak of the latent variable as if it were psychological, but it is unclear why they feel licensed in doing so. Further, if we assume (in accordance with the model) that the "different indicator values" regard N-indicators, then the claim is true

the formative model is only construed as determining the relationship between *measurements*, it only accounts for data, and does not explain phenomena.[5] The deficiency of this view, in terms of explanatory import, becomes clear in the authors' case studies. In one study, they obtain data correlating "intelligence" (defined as measurments on a series of tests) and brain volume (defined as measurements taken using voxel-based morphograpy, or VBM). They then apply both formative and reflective models to the data and observe that the formative model generates a considerably better fit in correlating these two data sets than the reflective model. The authors thus conclude that intelligence supervenes on brain matter and that the metaphysical relation between them is "solved" via the measurement. This conclusion suggests that the model completes the *investigation* by providing a satisfactory view of the metaphysics of intelligence. We contend that this is entirely insufficient as a neuroscientific explanation of intelligence, one that no cognitive neuroscientist should (or probably would) be satisfied with. The reason is that we have only accounted for data. Both the nature of intelligence and of the mechanisms underlying it remain a complete mystery. In the next section, we offer a very different view of how to import metaphysics usefully into cognitive neuroscience, which takes a metaphysical posit as a *beginning to* and a *guide toward* mechanistic explanation.

Identity Claims and Mechanistic Explanations in Cognitive Neuroscience

As we argued in the previous section, cognitive neuroscience is typically engaged in explaining phenomena. The form of such explanation is typically mechanistic. Recently philosophers of biology have attempted to articulate what mechanisms are and how they figure in explanations. Fundamentally, a mechanism consists of parts performing operations that are organized and coordinated so as to produce the phenomenon of interest; a mechanistic explanation characterizes the responsible mechanism and shows how it could produce the phenomenon (Bechtel & Abrahamsen, 2005; Bechtel & Richardson, 1993/2010; Glennan, 1996; Machamer, Darden, & Craver, 2000).[6] The quest for mechanisms has been pursued as well in psychology—the cognitive revolution can be construed as advancing the view that the mind is an information-processing mechanism (or a set of such mechanisms). For the most part, psychologists investigating human behavior were limited to behavioral tools such as reaction times and error patterns to assess proposed mechanisms. They lacked techniques for identifying the brain processes that were responsible for these operations in humans and so could not use information about brain parts and their connections to constrain their proposed mechanisms.[7] This situation changed dramatically with the introduction of techniques for employing positron emission tomography and, subsequently, magnetic resonance imaging to identify brain regions involved in cognitive activities in humans. In conducting this research, cognitive neuroscientists attempt to localize operations differentiated in psychological tasks in specific brain regions (using blood flow as a proxy for neural activity) and then employ knowledge gleaned from the brain to explain cognitive activities.

The first step in a mechanistic account is to delineate the phenomenon for which a mechanism is sought. Like every step in developing a mechanistic explanation, the delineation of a phenomenon is fallible and subject to revision. To relate the mechanism to the brain, it must be localized in some (possibly distributed) brain region. The second step is to decompose the mechanism into component operations and localize these in component parts of the brain region. Both of the localization steps involve identity claims, in which a psychological process (or subprocess) is taken to be identical to neural processes. However, these are a very different sort of identity claim than the ones the authors of the target article suggest. In particular, such identity claims are advanced to *initiate* inquiry and serve to guide subsequent research. To capture this aspect of identity claims in science, Bechtel and McCauley (1999; McCauley & Bechtel, 2001) refer to them as *heuristic* and refer to the resulting version of identity theory as *heuristic identity theory* (HIT). These identity claims are viewed as fallible, and one of the objectives of the research projects to which they give raise is to revise the identity claims as needed.

An identity relation is a particularly strong relation that embraces Leibniz's law of the indiscernability

---

only if two people's brains are measured with *the very same* styles of measurement. Add or subtract to the data and the latent variable changes. Assigning the same *numerical value* to the latent variables would then *not* be assigning the "same" value to the "same" latent variable; it would not be multiple realizability.

[5]In a way, the situation is worse for the formative models. Because the latent variable is only posited to *determine* the measurements, instead of *causing* them (as on the reflective model—see footnote 2), it is unclear that the formal rendering of the latent variable is invoking a metaphysical claim *at all*. This is related to the antirealism worry in the previous footnote.

[6]In most accounts, mental simulation is viewed as sufficient to show how the mechanism could generate the phenomenon being explained. But in many biological mechanisms, including ones found in neuroscience, the organization of operations is not sequential and the operations themselves are nonlinear. Accordingly, computational models and application of tools from dynamical systems theory are required to explain how the parts and operations work together to generate the phenomenon. Bechtel and Abrahamsen (2010) characterized such explanations as "dynamic mechanistic explanations."

[7]Research on the brains of nonhuman animals was possible, though, and yielded important insights into the mechanisms involved in vision (Bechtel, 2008), memory (Craver, 2007), and other phenomena.

of identicals: Everything true of the entity when it is denoted in psychological vocabulary must be true of it when denoted in neuroscientific vocabulary. For heuristic purposes this is vitally important, because any feature noted in either characterization that cannot be captured in the other characterization is a spur to revision.[8] To illustrate the heuristic role of heuristic identity claims in developing mechanistic explanations in cognitive neuroscience, we focus here on a case—the proposed localization of human face recognition to activation in a particular area of the fusiform gyrus—which has generated a research endeavor that is still in its early stages.

In a widely cited article, Kanwisher, McDermott, and Chun (1997), relying on differential activation measured in an fMRI study, advanced the strong hypothesis that an area in the fusiform gyrus, which they dubbed the "fusiform face area" (FFA), is a module for face recognition. The claim was also supported by neuropsychological data, in which lesions to the area produced symptoms akin to prosopagnosia (inability to recognize faces; Wada & Yamamoto, 2001; see Kanwisher & Yovel, 2006, for a discussion). The identity claim, then, is that the process of face recognition is *identical* to the activity of the FFA. The proposed identity was soon challenged. Gauthier and colleagues noticed that the FFA is also activated differentially by a variety of stimuli, including bird, cars, sculptures, and facelike figures that they call "Greebles." Moreover, they found an effect of training on activation in the FFA—activation increased with further exposure to Greebles (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999)—and showed that activation was higher for cars and birds if the subjects were experts in those subjects (Gauthier et al., 2000). This led Gauthier and colleagues to argue that the FFA is not a face area but a more general "expertise" area, involved in categorizing objects with which one is familiar. From the viewpoint of HIT, this alternate hypothesis clearly represents a revision of the original identity statement. Given that identicals are indiscernible, the *process* of face recognition cannot be identical to the activation of the FFA, because they have different observable properties—activation of the FFA occurs in response to cars, but face recognition does not.[9]

The important thing to notice is how the field has developed in the few years since this controversy emerged. One strategy has been to tease apart the specific contributions of the FFA in relation to the more general object-recognizing features of the ventral visual pathway, of which it is a part (Grill-Spector, Golarai, & Gabrieli, 2008). It is important to note that attempts have also been made to uncover functional subregions within the FFA. Both Grill-Spector, Sayres, and Ress (2006) and Haxby (2006) used high-resolution fMRI to locate particular, interspersed areas within the FFA that are preferentially activated to faces as well as ones that responded preferentially to other categories. Grill-Spector et al. claimed that the FFA is in fact "reliably heterogeneous" as to its structure and function (p. 1177), whereas Haxby suggested that, at a fine-grained level, function in the FFA is "distributed" across several independent subunits. Both views suggest that the original FFA data were the result of averaging over activation from several distinct subunits within the FFA.[10] From the standpoint of mechanistic research, this development represents an early attempt at *decomposing* the FFA. After a rough localization of facial recognition to the FFA (through the first identity claim), research is now progressing (through *revisions* to the first identity claim) to discover the fine-grained mechanisms and subprocesses involved in producing the phenomenon.

The research just discussed relies on manipulating sensory inputs and recording changes in the responses of various brain regions. A very different strategy is to manipulate a component of a proposed mechanism and determine the consequences of the manipulation on behavior. Pitcher, Walsh, Yovel, and Duchaine

[8]Supervenience is a much weaker relation than identity, only requiring that operations characterized in neuroscience terms always map onto the same psychologically characterized process but not vice versa. As noted by Kievit et al., supervenience is compatible with the psychological process being multiply realized. Putnam (1967) argued that mental states are multiply realizable, because different species appear to exhibit the same mental states despite major differences in their physical brains. Multiple realization is often taken as the death knell to identity claims. But, as Bechtel and Mundale (1999) argued, neuroscientific research has long treated activity in brains that exhibit morphological differences as the same. In doing this neuroscientists use a coarse-grained account of neural states that is comparable to the coarse-grained account required to treat the psychological states as the same. If one insists on finer-grained accounts in neuroscience so as to recognize differences in brain processes (as the multiple-realizability argument does), a comparably finer-grained psychological account will also find differences between species (or individuals). The important point to note here is that if the relation of mental states to brain states is to serve as a productive heuristic in guiding research, it is identity, not supervenience, that is required. There may be legitimate cases of multiple realization, resulting for example from convergent evolution, but the heuristic identity theory advocates that researchers settle for supervenience only if serious efforts fail to reveal psychological differences in brains that are differently organized.

[9]Here we gloss over several of the early debates that occurred in the field. Defenders of the modular position, citing the fact that activation of the FFA was slightly greater in response to faces than other stimuli, argued that the FFA is *primarily* a face recognition area and that the other activations were ancillary or residual (Kanwisher, 2000). This does not alter the situation from the standpoint of HIT, however—the FFA being *primarily involved* in face recognition is not the same as the claim that it is *identical* to face recognition. This move by early defenders of the modularity hypothesis, then, was in fact an early recognition of the need for further research, as we illustrate next.

[10]For a differing opinion based on a technical objection to the use of high-resolution fMRI, see Baker, Hutchinson, and Kanwisher (2007).

(2007) pursued such a strategy, applying repetitive inhibitory trans-cranial magnetic stimulation (TMS) to the right occipital gyrus, an area prior to the FFA in the ventral stream that had itself previously been implicated in face processing. When TMS was applied, facial discrimination was severely limited, thereby suggesting a multistage model of facial processing, in which activation in the occipital gyrus creates an "initial representation" (Pitcher et al., 2007, p. 1569) of the face. This result is relevant to both strategies just discussed: the further localizing of *specified* functioning (i.e., initial face representation vs. face recognition simpliciter) to different brain areas represents a more fine-grained account of the phenomenon, and a new spur to further decomposition.

Ideally, researchers would be able to apply the same approach to the structures within the occipital gyrus and FFA to further elucidate their contributions. However, there are limits in the spatial resolution of such techniques as TMS, and naturally occurring lesions rarely occur only within a functionally specified unit. The alternative is to seek model organisms on which more fine-grained recording techniques (in vivo electrophysiology) and manipulations (artificial lesions) can be employed. Model organisms have been vital to our understanding of both memory (mice) and vision (cats, macaques). Attempts are being made to find homologous processes between macaque and human facial processing (Tsao, Moeller, & Freiwald, 2008), but as yet these are still at the behavioral and coarse-structure level. Eventually developing the ability to perform fine-grained manipulations on subcomponents of the facial processing system, however, will be vital in understanding the mechanism. Computational models can also provide insight by making specific predictions about observable consequences of fine-grained processes without requiring actual manipulation; computational models have already been advanced to test the predictions of the expertise hypothesis (Palmeri & Gauthier, 2004). Although these strategies are in their early stages, both will undoubtedly prove important in advancing mechanistic understanding of facial processing.

This brief sketch of cognitive neuroscience research on face recognition suffices to illustrate the variety of methods cognitive neuroscientists commonly deploy in developing mechanistic explanations of cognitive phenomena. Unlike the psychometric strategy proposed by Kievet et al., these are aimed at revealing the mechanism responsible for a particular psychological phenomenon. This sketch also illustrates the role that proposing and revising identity claims plays in neuroscientific research. Kanwisher's hypothesized identity of face processing with activity in the FFA did not meet the conditions set by the indiscernibility of identicals for identity claims. Far from being a failure, however, the initial hypothesis has served as an important guide to developing more fine-grained theories, ones that both decomposed the FFA into smaller units and their corresponding processes and that connect the functioning of the FFA to other regions and their processes. Of vital importance is the fact that both the conceptions of the *phenomenon* (the dividing of facial processing into early and late stages) and of the underlying mechanisms (the switch from modular to distributed, and from univocal to substructured functional units) are modified in this process, with the developments in one constraining developments in the other. In each case the conceptions are modified from broader or more coarse-grained views of their targets to more specified, fine-grained views. We return briefly to the FFA example in the next section to illustrate the role of research at different levels within a mechanism.

Levels, Mechanisms, and Identity Claims

Mind and brain are commonly portrayed as occupying different levels in nature. If this were correct, mental and physical processes clearly could not be identical. So we must consider what is meant by a *level* in this context. Sometimes levels are identified in terms of inquiries, as when Oppenheim and Putnam (1958) spoke of disciplines as at levels, or when Marr (1982) differentiated computational, algorithmic, and implementational levels of analysis. Such accounts, however, do not pick out levels in nature, as the same item in nature could be analyzed in different disciplines or using different tools of analysis. A more ontologically oriented view of levels begins by noticing that there is a compositional, or part–whole, relation within many natural systems: in us, molecules are parts of cells, cells are parts of organs, organs are parts of us, and we are parts of societies. Some have tried to analyze such composition levels in terms of size (Churchland & Sejnowski, 1992), but if our reason for tracking levels is to facilitate our understanding of the mechanisms responsible for phenomena, a better approach is to turn directly to mechanisms and to count as one level the mechanism and the other entities (many themselves mechanisms) with which it interacts and as a different, lower level those parts with operations that are involved in the functioning of the mechanism (Craver, 2007).

On this construal, the investigation of a mechanism is by definition an interlevel pursuit: Decomposing a mechanism into its parts and operations is to descend to a lower level (and hence such inquiry is widely regarded as reductionistic). But the identity claims on which we are focusing are within a level. They relate two different descriptions as picking out either the same mechanism or the same component of a mechanism. Often they relate a structural and a functional characterization of a mechanism or one of its parts, but sometimes they relate two functional decompositions (e.g., when these have been developed with different tools). Insofar as they relate two different accounts of an entity at the same level, identity claims are not themselves reductionistic. They do, however, play an important role in reductionistic inquiry because they can facilitate further decomposition. Identifying a psychologically characterized process with the activity of a brain region makes possible a further inquiry into how that brain region is able to perform the psychological process by further decomposing it. Because at this point the psychological tools may have reached their limits of application, the further decomposition, both structurally and functionally, is in neuroscientific terms. To return to the face recognition example: The research described in the previous section only decomposed the brain into regions. Yet much of neuroscience is focused on neurons and molecular processes within them. To link research on brain regions to cell and molecular processes, it is necessary to determine how neuronal processes represent information. Much interest and compelling research has been devoted recently to the role of inhibitory interneurons in creating gamma frequency oscillations among groups of interconnected pyramidal neurons (Sejnowski & Paulsen, 2006), but these investigations are still a work in progress. If, ex hypothesi, gamma oscillations prove to be the relevant process for representing information at the neuronal network level, it will still remain to be seen how these oscillations are inculcated and employed in particular contexts. In the context of face recognition, researchers will need to determine how neuronal groups in particular subunits of the FFA interact with each other and groups of neurons elsewhere in the visual pathway to produce oscillating networks (or groups thereof) that represent faces.

The scenario just sketched suggests further reductionistic extensions of current face-recognition research, but often it is also necessary to integrate this research with that at higher levels. In addition to taking mechanisms apart, it is necessary to recompose them so as to show how they functions in their context. The term top-down causation is sometimes invoked to characterize how processes at high levels affect those at lower levels. Kievit et al. note Craver and Bechtel's (2007) discussion, which emphasized the differences between interlevel relations and causal relations. As normally construed, causes are independent of, precede, and produce their effects by some action. This is not true of interlevel relations—when a part of a mechanism is altered, the whole mechanism is thereby altered, and a mechanism cannot be altered without some part of it being altered. There is no action by which the part changes the whole. The two effects are not independent, and they occur simultaneously. When a causal factor impinges on a mechanism and thereby changes a component, a cascade of other changes may occur in the mechanism, with the consequence that the mechanism is further changed and interacts differently with the world. Strictly speaking, there is no bottom-up or top-down causation but only causation within levels. But by combining constitution relations between levels and causal relations at each level, one can capture the phenomena to which advocates of top-down causation draw attention. One can, moreover, ask (as Kievit et al. seek to do when posing the question at which level change occurred first) whether the relevant causal processes involved operations in the mechanism or between the mechanism and its environment. The crucial issue to be addressed here is the level at which the causation occurred, not the temporal order. Then experimental tools intervening on processes at specific levels can facilitate assessment of causes. What is critical here is that neither interlevel relations nor identity relations be construed causally.

The picture that is emerging from our discussion is what we might call one of "level-building." Phenomena at the original "psychological" level (face recognition) are decomposed into more fine-grained processes (early vs. late stage), which are identified with operations within specific brain areas (areas of the visual pathway, subunits of the FFA). Simultaneously, mechanisms at the cellular or molecular level (e.g., synchronization of network oscillations) are specified to account for their contributions to cognitive operations in particular contexts, such as the distributed activation of subunits in facial processing. As previously mentioned, the identity statements that end up succeeding are *intralevel*—they are between different descriptions of phenomena at the same level of complexity. Eventually, through the continued processes of decomposition and recomposition, we build to levels where, for instance, descriptions of neural operations involved in the subunits of the FFA have the same observable properties as the (suitably fine-grained) processes attributed to those units, thereby meeting the conditions for the indiscernibility of identicals. Thus, at the culmination of successful research, the metaphysical relation between the descriptions flanking the equals sign is unproblematic—there is no longer any "gap" to be overcome.

Conclusions

We have contrasted the characterization of psychological and neural processes outlined in the target article, which uses psychometric correlations between measures to identify latent variables, with the heuristic identity claims and ensuing mechanistic research, which we find to be common in cognitive neuroscience. We have explored, using research on face recognition, how such research spans levels within mechanisms but invokes identity claims only within levels. In the research on face recognition the discrepancies that force revisions to identity claims could be identified qualitatively. We fully anticipate, as research develops, that these assessments will become more quantitative. Thus, the contrast of our views with those of Kievet et al. turns not on whether identity claims are assessed qualitatively but whether they result from establishing relations between measures in the course of mechanistic explanation of phenomena.

Note

Address correspondence to William Bechtel, Department of Philosophy-0119, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0119. E-mail: bill@mechanism.ucsd.edu

References

Baker, C. I., Hutchison, T. L., & Kanwisher, N. (2007). Does the fusiform face area contain subregions highly selective for nonfaces? *Nature Neuroscience, 10*(1), 3–4.

Bechtel, W. (2008). *Mental mechanisms*. London, UK: Routledge.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441.

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A, 41,* 321–333.

Bechtel, W., & McCauley, R. N. (1999). Heuristic identity theory (or back to the future): The mind–body problem against the background of research strategies in cognitive neuroscience. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (pp. 67–72). Mahwah, NJ: Erlbaum.

Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science, 66*, 175–207.

Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Cambridge, MA: MIT Press. (Original work published 1993)

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review, 97*, 303–352.

Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspective, 6*(1), 25–53.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110,* 203–219.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Craver, C. F. (2007). *Explaining the brain: What a science of the mind–brain could be*. New York, NY: Oxford University Press.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547–563.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience, 2,* 568–573.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12,* 495–504.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 50–71.

Grill-Spector, K., Golarai, G., & Gabrieli, J. (2008). Developmental neuroimaging of the human ventral visual cortex. *Trends in Cognitive Sciences, 12,* 152–162.

Grill-Spector, K., Sayres, R., & Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature Neuroscience, 9,* 1177–1185.

Haxby, J. V. (2006). Fine structure in representations of faces and objects. *Nature Neuroscience, 9,* 1084–1086.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3,* 759–763.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361,* 2109–2128.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

Marr, D. C. (1982). *Vision: A computation investigation into the human representational system and processing of visual information*. San Francisco, CA: Freeman.

McCauley, R. N., & Bechtel, W. (2001). Explanatory pluralism and the heuristic identity theory. *Theory and Psychology, 11*, 736–760.

Oppenheim, P., & Putnam, H. (1958). The unity of science as a working hypothesis. In H. Feigl & G. Maxwell (Eds.), *Concepts, theories, and the mind–body problem* (pp. 3–36). Minneapolis, MN: University of Minnesota Press.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5,* 291–303.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology, 17,* 1568–1573.

Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind and religion* (pp. 37–48). Pittsburgh, PA: University of Pittsburgh Press.

Sejnowski, T. J., & Paulsen, O. (2006). Network oscillations: Emerging computational principles. *Journal of Neuroscience, 26,* 1673–1676.

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences, 105,* 19514–19519.

Wada, Y., & Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *Journal of Neurology, Neurosurgery & Psychiatry, 71,* 254–257.