dialectica

Vol 75 : 3
September 2021

The Legend of Hermann the Cognitive Neuroscientist

Bryce Gessell

doi:10.48106/dial.v75.i3.03

Bryce Gessell. 2021. "The Legend of Hermann the Cognitive Neuroscientist." *Dialectica* 75(3): 1–1. doi:10.48106/dial.v75.i3.03.

The Legend of Hermann the Cognitive Neuroscientist

Bryce Gessell

I tell the tale of Hermann, a cognitive neuroscientist with transcendental aspirations. Hermann's story illustrates the fundamental problem of cognitive ontologies, which is the problem of knowing whether the background conceptual scheme for our psychological theories is correct. I show why this problem is so fundamental, how it arises from the nature of psychology as a science, and why various current approaches to solving it are not likely to be successful. The problem, I argue, pushes us toward instrumentalism about mental concepts and categories, in both psychology and cognitive neuroscience.

1 The Legend

Once upon a time, there was a cognitive neuroscientist named Hermann. Like his colleagues, Hermann read articles, applied for funding, and was a proficient neuroimager. He taught classes and went to department meetings. But unlike his colleagues, Hermann harbored a dark secret. It was a secret blacker than the coffee he drank while explaining the Libet studies to his undergrads for the fiftieth time. He dared not reveal the secret to anyone, even on his many fake Twitter accounts, lest the information somehow be traced back to him: Hermann was a Kantian.

Graduate school had been difficult for Hermann. A convert to transcendental philosophy at age 17, he didn't share his classmates' enthusiasm for cutting-edge theories of mental processes. He just couldn't see the point of devising or testing newfangled psychological concepts like "attentional control" and "reward-prediction error". After all, hadn't Kant already outlined the *true* psychology back in the 1780s? What more could the world want?

But nearly failing his first psychology courses taught Hermann never to disclose his true convictions, and so he dutifully read his textbooks and reproduced the "correct" answers on tests. He asked questions at department talks to throw his supervisors off the trail. He gave papers on his lab's work at the APA and SPSP. Those results led to a dissertation on neuroeconomics, which he resented while writing and loathed after it got him a social neuroscience postdoc. Yet once again he did what he was supposed to, and scanned countless fMRI subjects while they watched videos of people talking and laughing. He always wrote his findings up on time and sent papers to well-targeted journals. Many were accepted, some even at prestigious venues.

In reality, though, Hermann was just biding his time. All through graduate school and his postdoc, Hermann pretended to believe in the constructs of contemporary psychology, but deep down, he was just waiting for the moment when he could follow his heart. At last, the years of hypocrisy and dissimulation paid off: his postdoc papers struck a chord with the right committees, his job talk had the perfect jokes ("Based on the work of my very warlike colleague, Sarah Bellum, we…"), and he dazzled the right group of faculty. Hermann landed two big grants and a tenure-track job. To celebrate he took a long walk down the lane near his house at precisely 3:30 pm. On the walk he contemplated his future and looked at sticks; he knew the real work was just beginning.

The next morning Hermann gathered his notes from countless magical nights with Immanuel and got his real research program underway. His goal was simple: find the neural correlates for all the major constructs in Kant's psychology. With his detailed knowledge of the first *Critique* and other texts, Hermann knew that his work would not involve conceptual difficulties. Methodology wasn't a problem either, since his grad-school education was more than sufficient. He was merely doing what every other cognitive neuroscientist did—he just happened to be doing it within a Kantian framework.

It was not hard to find the neural correlates for the faculty of judgment, for example. While inside the fMRI scanner, his subjects read and reflected on propositions. Hermann carefully counterbalanced the stimuli to control for effects like variable propositional content and emotional valence. To the surprise of no one, his statistical analyses showed that certain brain areas activated during the tasks. These crucial areas showed regular patterns both across subjects and across studies. In this way, he identified not only the cortical regions engaged in acts of judgment, but also the sub-regions which process the various logical forms of judgments. He connected judgments of quality to one area and judgments of relation to another, and judgments of quantity and modality to interconnected networks.

Hermann made similar discoveries about spatial representations, thereby illuminating the neural mechanisms of "the form of all appearances of outer sense." (Kant 1998, A26/B42) His work produced the first map of the Kantian cortex.

No, day-to-day research was not the hard part. The real problem was time. For his arguments to be persuasive, Hermann knew that he needed a lot of data, and needed the computing cluster to analyze it with fancy statistics. But he also knew that he couldn't let anyone find out what he was doing. So he stonewalled his colleagues when they asked about his results; he ignored emails from his department chair; he kept his unfortunate non-Kantian graduate students in the dark about the true import of their work. The tipping point came at his third-year tenure review. The neuroscience faculty was ready to give him marks for unsatisfactory progress, which would have been grounds for dismissal, but a letter of recommendation from Hermann's postdoc director saved the day. Her letter assured department members of Hermann's potential, promised that he would revolutionize the field, and urged them to retain him. Hermann barely survived the vote. He knew he needed to hurry—he had less than three years to go.

In the darkest times, when he doubted his life's work and his goal appeared most distant, Hermann comforted himself by reading what Kant once wrote to Samuel Thomas Soemmering. In a 1795 letter, Kant spoke to the anatomist Soemmering about the sense organs in the brain. Sensory representations had to be combined, Kant said, and it was incumbent on natural philosophers to "render that unity comprehensible by reference to the structure of the brain."(Kant 1999, 501) Hermann drew strength from his forebear's prescient understanding of his own research program, as he attempted to show the neural correlates for *a priori* contributions to cognition. Hermann also knew he carried on Soemmering's physiological work, to which Kant gave effusive praise, by finding the chemical mechanisms of the mental faculties.[1]

But the next three years passed, and since he released his results only in controlled trickles, Hermann simply did not publish enough to get tenure. It was understood in his department that he would have to leave after his seventh year. Then, at the start of that year, a miracle happened: stacks of finished manuscripts, all on the cognitive neuroscience of Kant's transcendental philosophy, appeared on the department chair's desk. All together, the manuscripts told a magnificent story of how the brain realized the posits

1 See Kant's preface to Soemmering's *On the Organ of the Soul* (Kant 2007, 222–226).

doi: 10.48106/dial.v75.i3.03

of Kant's psychological theory. Early papers laid the groundwork by finding brain areas for the most fundamental concepts, like the faculty of judgment, the forms of space and time, and the transcendental unity of apperception. Later work described connections among these concepts that even Kant had not noticed. Shorter papers filled in smaller details, and a single flagship paper—Hermann hoped to send it to *Neuron*—assembled the main results into a new, elegant, and powerful theory of the mind and brain. Always and everywhere, Hermann's results met and even exceeded accepted standards of experimental rigor and statistical significance.

At first, the chair was laughing as he leafed through the pile, imagining the pleasure he would feel at firing this Prussian charlatan. But the laughing stopped as he began to see the depth, creativity, and penetrating intuition with which Hermann had carried out his work. He convened a special faculty meeting to discuss the matter. On the one hand, Hermann had published nothing of note during his six years as assistant professor; on the other, he was now sitting on dozens of bold papers, each ready to submit. The chair asked the faculty for their opinions. "It's such a waste!" a recently-tenured associate professor yelled. "Seven years down the drain! This whole thing is a travesty, and a sham, and a mockery! I won't stand for it!" Many others agreed. But Hermann had his defenders, mostly among the older faculty. These full professors, now in the twilight of their careers, had seen countless psychological theories come and go. From their point of view, the conceptual framework of Hermann's research did not differ essentially from so many failed frameworks of the past.

In the end, Hermann's colleagues decided to give him a choice: he could either leave the university or back up his neuroscientific results with behavioral studies. The faculty supporting him were worried that Kant's view was too procrustean to be plausible in the modern age. They wanted to see behavioral results demonstrating that Kantian psychology could account for known complexities of human action. They did not think it could be done, but if Hermann were able to pull it off, they thought, they could not justify forcing him out.

Hermann felt he couldn't abandon his work now—not when he had come so far. So he designed an arc of behavioral studies to support Kant's psychology. Fortunately for him, behavioral results are faster and cheaper to get than neuroimaging, and Hermann Turk'd almost everything. In what became his *annus mirabilis*, Hermann completed his entire suite of studies, performed some requisite follow-ups, and wrote all his results before the end of the spring semester. He even made original discoveries about the structure of cognition from a Kantian perspective (these he considered submitting to philosophy journals, but seriously, what's the point?). Once again, the chair showed up to work one day to find another pile of papers on his desk, showing how to implement Kantian psychology to describe all aspects of human behavior.

He convened a second meeting, and for a second time, the question divided the members of Hermann's department. Some continued to think that Kantian psychology was unworkable in principle, and that the idea of a "Kantian cognitive neuroscience" was a farce. Others felt that Hermann's body of work was, in many respects, comparable to that of other faculty that the department had tenured. But all agreed on what Hermann had set before them: a coherent, exhaustive, and radical alternative to the contemporary conceptual framework of cognitive neuroscience.

The behavioral work showed Kantian psychological concepts to be much more flexible than anyone had realized, and sufficient to account for human perception, action, and memory. The evidence for the Kantian constructs was every bit as relevant and rigorous as it was for anything else in psychology. The imaging work showed, moreover, that these concepts had clear and reliable neural correlates, and that multivariate analyses could predict their instantiation in a wide variety of tasks. Indeed, it was not a matter of weighing evidence at all, for the evidence was equal on both sides—Hermann's brain data and behavioral studies were beyond reproach. Nor was it that Hermann had shown how to make certain Kantian constructs work *within* contemporary psychology. Rather, he was in fact offering a complete *replacement* for *all* of contemporary psychology. This Hermann's colleagues understood, and it was the root of their complaint. Hermann's work formed a complete science of human behavior which was fundamentally incompatible with competing approaches—that is, with *their* approaches. And he did it all with a brilliance and Teutonic flair that no one had ever noticed in him.

Department members faced a stark choice: dismiss an apparent rising star ("einen aufgehenden Stern", joked an older faculty member no one liked), or tenure a Kantian. They abhorred both options. On the one hand, they could get rid of him. But doing so would be an indictment on their own careers, for Hermann had done everything they had, and just as well, only with a different set of cognitive concepts. They realized that had the history of psychology gone differently, they might have been Kantians too. On the other hand, granting him tenure would sanction Hermann's revival of transcendental psychology—and let's face it, no one wanted that. The empirical evidence was

doi: 10.48106/dial.v75.i3.03

equal on both sides, and the practical consequences were all bad. How could they decide?

In the end, however, Hermann spared them the trouble. Having been asked to speak in his own defense, he instead offered his resignation. The *annus mirabilis* had ironed out the last wrinkle in his work, he explained, and so he had achieved his goal. There was nothing left for him to do. His results were just as good as theirs or anyone else's—he knew it, they knew it, and he knew they knew it. Hermann rose from his chair, grabbed a few of the big cookies they always had at faculty meetings, and walked out into the sunset. No one ever heard from him again.

Thus the legend of Hermann, the Kantian cognitive neuroscientist, was born.

2 The Moral

Hermann's legend makes several important points about cognitive neuroscience. I'll elaborate on some of them here as well as on the philosophical issues involved. I'll also consider some objections to my framing and conclusions.

To begin, I am not the first to tell a tale like Hermann's. Bub (2000) gave a version of it using phrenology, and so did Poldrack (2010). Others have told it as well (Uttal 2001; Anderson 2015).

All these versions involve *cognitive ontologies*. A cognitive ontology is the set of entities, processes, and constructs in one's theory of cognition.[2] We should understand "cognition" broadly here, as including sensation, perception, consciousness, and any other mental process or phenomenon. So if our ability to remember a phone number by silent rehearsal requires a "phonological loop" (Baddeley and Hitch 1974), then the phonological loop belongs in our cognitive ontology.

Most disputes in psychology concern the details of a cognitive ontology: whether this or that entity belongs in it, or whether some entity has this or that property. Memory researchers debate, for example, whether consolidation is distinct from *re*consolidation (Alberini and LeDoux 2013). Consolidation occurs when a memory becomes insensitive to disruption or change. But each time someone reactivates a memory, it becomes susceptible to interference

2 See Poldrack (2010) and Janssen, Klein and Slors (2017) for similar definitions. Anderson (2015) uses the term "taxonomy".

again. Is this latter event also just consolidation, or is it a separate process with different temporal and mechanistic profiles (Lee, Nader and Schiller 2017)? Our answers to these questions determine part of our cognitive ontology, and we can ask similar questions across psychology.

In turn, most research programs in cognitive neuroscience deal with mappings between a cognitive ontology and brain structures. The mappings involve local questions about processes like reconsolidation, but also global ones about which neural structure types we should map to. Philosophical theories about mechanisms (Piccinini and Craver 2011) and large-scale data projects (Yarkoni et al. 2011) try to solve these problems.

The workflow of a typical research program in cognitive neuroscience begins with whatever constructs the currently accepted cognitive ontology contains. Researchers then design tasks that they believe will involve those constructs. Next, they have study participants perform the tasks while some recording technique, such as fMRI or EEG, measures their neural activity. The hope is to find activity that exceeds a certain threshold or survives some correction for multiple comparisons. Should they find it, researchers map the construct they started with to the area showing the activity. They can then claim that the construct "engages" or "recruits" neural activity in that area. If they are careful, they will condition their claims on the tasks used, for tasks are inescapable mediators of mappings between mind and brain.

The legend of Hermann, however, is not about projects such as these. It is not about local disputes in psychology, nor the details of some mind-brain mapping. Rather, it is about which cognitive ontology we should prefer at the *general level*. It questions why a research program in neuroscience should begin with constructs from the received ontology of contemporary psychology at all. Why not select from an altogether different cognitive ontology? The history of psychology offers many choices. Hermann's tenure case also raises the possibility of the wholesale *replacement* of one cognitive ontology by another, where the replacing set of concepts is different from and even incompatible with the one replaced.

In short, the primary point behind Hermann's legend concerns what I call the *fundamental problem of cognitive ontologies*. Most studies in psychology don't touch this problem, for they work within an accepted ontology in order to refine it or fill in the details. The fundamental problem of cognitive ontologies is whether we should actually accept the received ontology, or prefer some other. Sure, a budding psychologist has practical reasons to reject Franz Joseph Gall's phrenological concepts as she begins her career. Chief

doi: 10.48106/dial.v75.i3.03

among them is that she'll never get a job by studying things like "veneration" or "amativeness". However, her practical reasons do not solve the in-principle problem of choosing a cognitive ontology to begin with. She could start her research just as well with the constructs of Aristotle, Galen, Christian Wolff, or anyone else with a theory of mind.

We can also put it this way: the fundamental problem of cognitive ontologies is knowing whether the conceptual scheme structuring your ontology is the right one. The problem is determining whether you have the correct conceptual language *in general*, not just in particular cases.

Kant's psychology is one such conceptual language. So why not be like Hermann and adopt it, instead of contemporary cognitive science, as the scheme to structure our whole ontology? Instead of "consciousness" we could talk about the "transcendental unity of apperception", for example. Kant wrote, "[t]he transcendental unity of apperception is that unity through which all of the manifold given in an intuition is united in a concept of the object" (Kant 1998, B139).

Assuming this is true, we can imagine various ways in which the unity of apperception might break down. People with akinetopsia or motion blindness do not have smooth perceptions of motion—their visual experience of motion is frame-by-frame, as it were, with no perceived connection between the frames. A good Kantian hypothesis would be that akinetopsia results from failing to properly combine the sensible data in the manifold. We could study this phenomenon in many ways: we could get behavioral profiles of people with akinetopsia-like symptoms and correlate our findings with life histories (Ovsiew 2014); we could test lesion patients with similar deficits (Rizzo, Nawrot and Zihl 1995); we could try to induce akinetopsia via transcranial magnetic stimulation and disrupt normal apperception ourselves (Beckers and Hömberg 1992). There would be many other avenues to explore. Some will scoff at this suggestion, but the point is that I have just described a research arc that would carry someone to associate professor and beyond. The published results would look an awful lot like psychology papers now, except Kantian concepts and a Kantian cognitive ontology would structure them.

I could provide more examples to deepen the point, or outline fMRI studies that Hermann could have done to plumb the implementation of Kant's psychology. But the actual history of psychology furnishes us with more and more plausible examples than we could ever hope to invent. The cycle of theory-replacement in the history of psychology *is* the existence proof for an in-principle problem.

There is a temptation to believe that, because psychology is a "science" now, its current cognitive ontology must stand on firmer ground than past ones. Can't we now draw sharper distinctions between different systems of memory? Don't we have better information about exact temporal profiles? Aren't we able to see better how entities in the ontology relate to each other? Yes, psychology does all this now, and it didn't or even couldn't do it in Kant's day. But we should not therefore infer that the accepted ontology has better epistemic credentials. The reason that items in our cognitive ontology have those properties is just that we now do psychology in a way that encourages us to identify those properties. Had we been doing psychology in Germany in 1800, but with modern methods, we could have discovered the same "facts" about the posits of Wolffian and Kantian psychology. That we could identify those properties, however, says little about their reality.

As such, there is no doubt that a real-life Hermann would succeed in finding neural correlates for the faculty of judgment, as described in the legend. He would have no trouble finding consistent, statistically significant patterns. His studies could use classic psychological testing methods like additive factors and subtraction. These methods work regardless of the entities in our cognitive ontology. They are varieties of experimental and task design, and any "justification" they confer on gathered data is irrespective of that ontology.

The problem of cognitive ontologies does not emerge because of modern methods, though two other (independent) methodological issues exacerbate it. The first begins in psychology: it is not difficult to find significant results in human cognitive and behavioral testing. Human behavior is amenable to description by many conceptual languages, which is why the history of psychology is so rich with ideas. A part of the problem stems from current experimental techniques, but another part is more endemic to psychological practice (Meehl 1967). The second methodological problem comes from neuroscience. Brains will show neural activations to anything and everything, so the fact that we have found an activation is not in itself very remarkable. We couldn't *not* have found an activation.

I will say a bit more about these two problems below, but they are not my primary concern. The legend of Hermann itself just illustrates the fundamental problem of cognitive ontologies and some associated philosophical issues. What, then, should we do about it?

Given the nature of psychology, I think the right move is to be instrumentalist about psychological theories. Earlier I spoke about the "right" ontology,

doi: 10.48106/dial.v75.i3.03

and finding the "correct" conceptual language. Human behavior and its neural basis may not be the kind of phenomena that allow true theories; it may just be that certain ontologies are better for certain situations. We could do cognitive neuroscience with one of many ontologies, but we pick the one that seems most useful for our purposes, whatever those may be.[3]

Not all practitioners of the mind-brain sciences want to go instrumentalist, however. Other ways to respond seek to carve out more room for realism and a "correct" ontology. Let's look at some of them.

Adapting Anderson's (2015) discussion of mind-brain mappings, we can distinguish three realist-motivated approaches to the fundamental problem of cognitive ontologies. The first, taken by the vast majority of psychologists and cognitive neuroscientists, is the *conservative* approach (Price and Friston 2005). This attitude assumes that the correct conceptual scheme is probably a lot like the one we have now, and so our cognitive ontology only requires local tweaking. The second approach is *moderate*. It attempts to let the brain decide which of two cognitive constructs is better. The third is the *radical* approach. It suggests a re-thinking of "the very foundations of psychology in light of evidence from neuroscience and evolutionary biology" (Anderson 2015, 70).

None of these three approaches to the challenge of cognitive ontologies necessitates realist commitments, though all three trend in that direction. All three suggest that there is a "true" ontology and that either we've already found most of it, or we at least know the way to get there. I'll discuss each approach in more detail below, and then explain why I don't find them very promising.

The first approach is conservative. It suggests that we already have most of the pieces for a true cognitive ontology—they're just the constructs of contemporary psychology. This approach takes the apparent success of psychological science as evidence of the truth of its claims, and since those claims involve elements in an ontology, the elements must therefore exist.

The problem with the conservative response is that it begs the question against someone like Hermann. Hermann suggests replacing the current ontology with another one; to say we can't do that, because the one we have now is true, assumes what Hermann denies.

It's also wrong to think that the "success" of psychological science, or the fact that each published paper finds an effect, creates a problem for Hermann's

3 This view shares something in common with the position outlined in Francken and Slors (2014).

Kantian view. Citing particular successful studies or even batches of them does not support conservatism. This is because the evidence for this or that current psychological theory is not thereby evidence for the background conceptual scheme in which those theories are framed and tested. As noted above, psychology is such that we cannot help but find evidence for virtually any construct we go looking for (Meehl 1967; Open Science Collaboration 2015). Thus finding evidence for some process says very little about the truth of the conceptual language describing that process. In other words, the reason we don't have empirical evidence for Kant's psychology is simply that no one has bothered to gather it yet. If a real-life Hermann ever comes around, he'll find all the evidence he could want, but he'd be no closer to establishing the reality of the Kantian cognitive ontology.

The second approach to the problem of cognitive ontologies is moderate. It uses brain data to adjudicate between competing or incompatible psychological constructs, thus letting the brain "speak for itself". The brain can do this in various ways. One is when competing cognitive categories make different predictions about their neural correlates. We can test these predictions by measuring brain activity during task conditions that involve the categories. Another way is through multivariate analyses, which use patterns of neural activations to predict cognitive constructs or representational categories of stimuli.

The moderate approach faces several challenges. For one, while brain data might be useful for comparisons between constructs, it cannot give an absolute measure of a construct's reality. This point leads to a more serious problem, which is that even brain data cannot adjudicate between entire conceptual schemes or whole cognitive ontologies. Indeed, the brain is a fit counterpart for psychology: it will always give us *some* evidence of whatever we test for. Bub (2000) and Poldrack (2010) used phrenology in their version of Hermann's tale because there is no question that phrenologists, had they used fMRI, would have found copious activations strongly correlated to their phrenological categories, and strongly predictive of those categories in multivariate studies. The same is true for Hermann's transcendental concepts, and for any other set of concepts we care to check: no matter what they are, we will find some neural signature of them—*but it does not follow that they are real*. Brain "data" or "evidence" usually aren't evidence for the reality of the mental construct being tested. This point seems to be either ignored or misunderstood by many philosophers and scientists.

doi: 10.48106/dial.v75.i3.03

Another way of putting the issue is to say that, while the moderate approach wishes to let the brain speak for itself, our neural organ can really only do so in a language that *we already understand*, where "we" are the designers and interpreters of experiments. If brain data is to shed light on human thought or behavior, we must interpret that data using cognitive concepts. Even the simplest interpretations therefore rely on entities in a cognitive ontology, even when those entities appear to be mere folk-psychological categories like perception, belief, or desire. Those basic categories also inform experiment and task design, as researchers use folk psychology to reach broad (albeit general) agreement on how psychological constructs, tasks, and experimental conditions relate.[4] That whole psychological apparatus forms a conceptual scheme for studying the mind and brain.

But if we bring to the brain a language we already understand—a worked-out cognitive ontology—then the moderate approach begs the question against Hermann no less than the conservative approach does. This criticism also applies to ontology construction if the analysis uses previously existing cognitive constructs to structure the data; such analyses comprise the majority of "data-driven" methods (Poldrack 2010; Yarkoni et al. 2011; Yeo et al. 2015; Tamar et al. 2016; Eisenberg et al. 2019; Genon et al. 2018; Bolt et al. 2020).

The third and final approach is the radical one. My objections to the first two approaches suggest that Hermann himself begs the question against current cognitive science since he brought a worked-out cognitive ontology of his own to studying the brain. But there are even more radical approaches that try to avoid begging the question. One example is Cisek (2019), who synthesizes a new cognitive ontology by analyzing the evolutionary history of simple behavioral systems. Another attempt is Pessoa, Medina and Desfilis (2021), who reject "standard mental terms" and instead found a new cognitive ontology with "complex, naturalistic behaviors".

It's too early to know whether projects like these will succeed. If a "true" cognitive ontology exists, these are our best bets to find it, because they throw out our current conceptual language and start with the evolutionary environment. There are other radical approaches that I think we can object to, however, so I will focus on those.

Other examples of the radical approach to cognitive ontologies use large data sets to find non-obvious dimensions or axes in brain activations. Call this the "latent structure" strategy (Yarkoni et al. 2011). I'll discuss the strategy

4 I thank a reviewer for making this connection clear.

a bit and then present a problem for it, which applies in varying degrees to other radical approaches.

The latent structure strategy uses computational techniques to find structure in neural data. The assumption is that the data's latent dimensions may trace the contours of categories the brain itself uses to organize cognition. In this approach, the brain goes beyond playing arbiter for competing constructs to reveal a brand-new set of categories. For example, Chen et al. (2017) use independent component analysis (ICA) with resting-state fMRI data from hundreds of scans to identify four previously hidden brain networks. The authors dub them the "auditory", "control", "default mode", and "visual" networks. Biswal, Mennes and Xi-Nian Zuo (2010) perform a similar analysis on resting state data, and Schaefer et al. (2018) use functional connectivity to produce a new cortical parcellation.

These analyses outdo Hermann's because they are based purely on brain measurements. You apply a technique like ICA and a robust structure emerges that may have been impossible to detect otherwise. Unlike every other approach, you need not bring anything to the table other than the data. Prior to identifying the structural contours, no part of any background conceptual scheme plays a role. This is another radical way of tackling the fundamental problem of cognitive ontologies, and perhaps another hope to avoid begging the question.

The challenge for latent structure strategies is *interpreting* what they find. Sure, Chen et al. (2017) find four separable networks. But where do the "auditory", "control", "default mode", and "visual" labels come from? Why interpret the networks with that conceptual language, instead of some other?

Now, the source for the labels is, of course, the authors' prior knowledge of similar networks. Chen et al. (2017) know that, in previous studies, participants who engaged in tasks requiring cognitive control showed activation patterns matching one of the networks they discovered. The authors then import those labels—those entities in the background cognitive ontology—into their own study, and use them to interpret the data. So even though the data's structure is *discovered* ontology-free, it can only be *interpreted* by some existing ontology or conceptual scheme. Just as we saw with the moderate approach, the brain can only speak in a language we already understand. The lesson is that big data may help introduce new neural categories, but it doesn't and can't provide the psychological labels for those categories.

Jerry Fodor and Ernie Lepore (1992, 1996) once developed a similar objection to Paul Churchland's semantic theory. Churchland (1989, 1998) devel-

doi: 10.48106/dial.v75.i3.03

oped a theory of meaning in which different aspects of conceptual content were represented by different dimensions in a high-dimensional neuronal activation space. So, to use a simplified example, the concept "dog" might be represented by neural activations along dimensions like "furriness", "barking-ness", "four-footed", and so on. Various ranges of those dimensions define a high-dimensional solid that constitutes the concept "dog".

The crux of Fodor and Lepore's objection is that Churchland begs the question about the labels on the dimensions. Why does the first dimension in the activation space represent "furriness" instead of "barking-ness", or something else entirely? By taking the labels for granted, Churchland smuggles semantic terms into a theory that is supposed to explain how there could be semantics in the first place.

Latent structure strategies make the same mistake. Why is this particular structure the "control" network, and that structure the "default mode" network? Labeling the networks requires interpreting the data, but interpretation only happens through cognitive concepts we already have. In trying to discover the brain's categories for cognition, we smuggle in the psychological labels, and so accomplish nothing other than putting old wine into new bottles.

In sum, I see the fundamental problem of cognitive ontologies as leading us toward instrumentalism about psychology. Although there are realist-friendly responses to this problem, most of them take the items in their cognitive ontology for granted, and we can't yet evaluate the ones that don't.

The moral of Hermann's legend is the problem I've been discussing, which connects to many issues in the philosophy of mind and of various sciences. Other than inertia and the vicissitudes of history, we have much less reason than we like to believe to prefer current cognitive ontologies over possible alternatives. And, as Bub (2000) notes, without some resolution for this problem,

> [we cannot] differentiate what is currently undertaken [in cognitive neuroscience] from a pointless activity in which inevitable differences between experimental and baseline conditions are falsely attributed specific cognitive interpretations that do not in fact correspond to reality (Bub 2000, 470).

I conclude by considering some objections to my arguments and the way I've set them up. First, you might say that this is all just a problem of reverse inference. Suppose my neuroimaging study discovers activation in brain area *X*. From previous studies, I know that *X* is associated with emotion, and so I infer that my subjects used emotional processing in my task, even though the task didn't explicitly involve emotion. This pattern of reasoning is called a reverse inference (Poldrack 2006). Reverse inferences require caution because area *X* could be involved in many other cognitive processes, not just emotion.

The problem of cognitive ontologies is not one of reverse inference, however. Reverse inferences have to do with evidence, and gathering more evidence alone does nothing to solve the problem. We have an enormous amount of papers published in cognitive psychology, but the sheer number does not resolve the in-principle problem of ontology selection.

A second objection could be that we could solve the problem with multivariate analyses in neuroscience. Both philosophers and neuroscientists sometimes believe that multivariate pattern analysis (MVPA), representational similarity analysis (RSA), and other multivariate techniques yield some special insight into brain function that ordinary univariate imaging analyses cannot (Nathan and Del Pinal 2017). I am skeptical of that view, but even if it were true, it would be irrelevant to my arguments. The problem of cognitive ontologies is not a methodological one—at least, not one internal to psychology or cognitive neuroscience as they are currently constituted. As I said above, certain methodological issues do exacerbate the problem, such as the ease with which we find publishable results in the mind-brain sciences. But it is not the current methods of psychology and cognitive neuroscience that give rise to the problem. It goes beyond the conceptual boundaries of either field and so we cannot solve it with more sophisticated statistics.

Someone might also object that the problem of cognitive ontologies is really an issue of underdetermination of theory by data (Aktunc 2021). According to this objection, alternative ontologies only look like live options because we don't yet have enough evidence for our current one. But this objection also says that psychological theories are *theories*, and as such, they will always go beyond the data. Every theory in every science outstrips the available observations, and it's unfair to expect a cognitive ontology to be an exception. This objection can therefore say that the problem of cognitive ontologies is not an issue of principle; it's just the expected result of humans doing psychological science.

This objection is a sophisticated one. To lay out and respond to all the issues involved would take another paper. Here I will just give some reasons to think

doi: 10.48106/dial.v75.i3.03

that the problem of cognitive ontologies goes beyond the underdetermination of theory by data.

As we've seen, Hermann wasn't going to convince anyone of Kant's psychology, no matter how much his evidence "determined" his theory. While Hermann's work isn't real, the cycle of theory replacement in the history of psychology is, and we have no reason to think that the cycle will stop with something like our current cognitive ontology. Superficial similarities between psychology and other sciences, such as that they are practiced in universities and use quantified measurements and mathematical analyses, give the impression that psychology, like physics or chemistry, trods a monotonic path up the mountain of truth. But those similarities belie deep conceptual and interpretational problems which may be inevitable not only in psychology but also in the phenomena it studies.

In describing human behavior and mentality, we face a situation in which many distinct but mutually incompatible conceptual schemes could do the job. It isn't just the history of psychology that shows this; current cross-cultural psychology does too. Take "indigenous" or "local" psychological theories, which describe human thought and behavior in specific cultural contexts (Allwood and Berry 2006). Rather than fitting received psychological categories to non-Western peoples, indigenous psychologies develop new categories tailored to their environment. Inputs to this development include literature, observations of behavior, self-reports, and past scientific evidence (Cheung et al. 1996). The results are psychological theories that may account for patterns of thought and behavior better than traditional (Western) theories.

One of the most empirically successful indigenous psychologies is the Chinese Personality Assessment Inventory, now known as the Cross-cultural Personality Assessment Inventory (CPAI). In addition to categories from the standard five-factor personality model, the CPAI includes psychological constructs like "Harmony", "Ren Qing" (relationship orientation), "Ah-Q Mentality" (defensiveness), and "Face" (Cheung et al. 2001). These constructs constitute a personality factor, "Interpersonal Relatedness", which is not reducible to other personality theories (Cheung et al. 2003).

If "Interpersonal Relatedness" and associated constructs like "Ren Qing" and "Ah-Q Mentality" are incompatible with other psychological theories, then what do we say about the state of the science? Underdetermination suggests that we're just lacking the evidence to decide between them, whether or not psychology is capable of providing it. But it's not a leap to think there may be some real indeterminacy here, and that there simply is no fact about whether "Ren Qing" is real. We can study it, we can use it, and we can endorse it, but we don't need to conclude it must exist.

There are indefinitely many conceptual schemes for psychology, limited only by our imagination. Whatever they are like, the brain will oblige with consistent profiles of activation. If the data underdetermines all the available theories to the same degree, then maybe the problem lies not in our ability to gather evidence but in the *Dinge an sich*.

One final objection. In a "no-miracles" spirit, one may say that our current ontology can't be *that* wrong, since psychology and neuroscience are so successful. To those with the courage to make this response: I envy your faith, but see no reason to share it.*

Bryce Gessell
0000-0003-4424-5627
Southern Virginia University
bryce.gessell@svu.edu

References

Aktunc, M. Emrah. 2021. "Productive Theory-Ladenness in fMRI." *Synthese* 198(3): 7987–8003, doi:10.1007/s11229-019-02125-9.

Alberini, Cristina M. and LeDoux, Joseph E. 2013. "Memory Reconsolidation." *Current Biology* 23(17): R746–R750, doi:10.1016/j.cub.2013.06.046.

Allwood, Carl Martin and Berry, John W. 2006. "Origins and Development of Indigenous Psychologies: An International Analysis ." *International Journal of Psychology* 41(4): 243–268, doi:10.1080/00207590544000013.

Anderson, Michael L. 2015. "Mining the Brain for a New Taxonomy of the Mind." *Philosophy Compass* 10(1): 68–77, doi:10.1111/phc3.12155.

Baddeley, Alan D. and Hitch, Graham. 1974. "Working Memory." *Psychology of Learning and Motivation* 8: 47–89, doi:10.1016/S0079-7421(08)60452-1.

* This paper began as a meme and has turned into something much more serious, mostly thanks to the members of the Imagination and Modal Cognition Lab and the SPACE lab at Duke. In particular, I thank the PIs of those two labs, Felipe De Brigard and Jenni Groh, for their support over so many years. Thank you to Derek Haderlie for his help. I would also like to thank an anonymous reviewer from another journal, and the various anonymous reviewers at *Dialectica*, who helped me make major improvements to the paper (and were willing to review such a bonkers paper in the first place). Most of all, I want to thank all my students at SVU—and especially those in my Minds, Brains, and Neuroscience course for spring semester 2022. To Nate, Nick, Hannah, Caroline, Drake, Allie, Elizabeth, Madelyn, Brett, Rachel, Katie, Caleb, Sabra, Gunner, Isaiah, Carter, Tay, Clayton, and Tyson: you're the best—and if this doesn't convince you, nothing will!

BECKERS, Gabriel J. L. and HÖMBERG, Volker. 1992. "Cerebral Visual Motion Blindness: Transitory Akinetopsia Induced by Transcranial Magnetic Stimulation of Human Area v5." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 249(1325): 173–178, doi:10.1098/rspb.1992.0100.

BISWAL, Bharat B., MENNES, Maarten and XI-NIAN ZUO, Clare Kelly, Suril Gohel. 2010. "Toward Discovery Science of Human Brain Function." *Proceedings of the National Academy of Sciences of the U.S.A.* 107(10): 4734–4739, doi:10.1073/pnas.0911855107.

BOLT, Taylor, NOMI, Jason S., ARENS, Rachel, VIJ, Shruti G., RIEDEL, Michael, SALO, Taylor, LAIRD, Angela R., EICKHOFF, Simon B. and UDDIN, Lucina Q. 2020. "Ontological Dimensions of Cognitive-Neural Mappings." *Neuroinformatics* 18(3): 451–463, doi:10.1007/s12021-020-09454-y.

BUB, Daniel N. 2000. "Methodological Issues Confronting PET and fMRI Studies of Cognitive Function." *Cognitive Neuropsychology* 17(5): 467–484, doi:10.1080/026432900410793.

CHEN, Shaojie, HUANG, Lei, QIU, Huitong, NEBEL, Mary Beth, MOSTOFSKY, Steward H., PEKAR, James J., LINDQUIST, Martin A., EOLOYAN, Ani and CALFO, Brian S. 2017. "Parallel Group Independent Component Analysis for Massive fMRI Data Sets." *PLoS One* 12(3), doi:10.1371/journal.pone.0173496.

CHEUNG, Fanny M., CHEUNG, Shu Fai, WADA, Sayuri and ZHANG, Jianxin. 2003. "Indigenous Measures of Personality Assessment in Asian Countries: A Review." *Psychological Assessment* 15(3): 280–289, doi:10.1037/1040-3590.15.3.280.

CHEUNG, Fanny M., LEUNG, Kwok, FAN, Ruth M., SONG, Wei-Zheng, ZHANG, Jian-Xin and ZHANG, Jian-Ping. 1996. "Development of the Chinese Personality Assessment Inventory ." *Journal of Cross-Cultural Psychology* 27(2): 181–199, doi:10.1177/0022022196272003.

CHEUNG, Fanny M., LEUNG, Kwok, ZHANG, Jian-Xin, SUN, Hai-Fa, GAN, Yi-Qun, SONG, Wei-Zhen and XIE, Dong. 2001. "Indigenous Chinese Personality Constructs: Is the Five-Factor Model Complete?" *Journal of Cross-Cultural Psychology* 32(4): 407–433, doi:10.1177/0022022101032004003.

CHURCHLAND, Paul M. 1989. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, Massachusetts: The MIT Press.

—. 1998. "Conceptual Similarity Across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered." *The Journal of Philosophy* 95(1): 5–32, doi:10.2307/2564566.

CISEK, Paul. 2019. "Resynthesizing Behavior Through Phylogenetic Refinement ." *Attention, Perception, & Psychophysics* 81(7): 2265–2287, doi:10.3758/s13414-019-01760-1.

EISENBERG, Ian W., BISSETT, Patrick G., ENKAVI, A. Zeynep, LI, Jamie, MACKINNON, David P., MARSCH, Lisa A. and POLDRACK, Russell A. 2019. "Uncovering the

Structure of Self-Regulation through Data-Driven Ontology Discovery .” *Nature Communications* 10(1): 2319, doi:10.1038/s41467-019-10301-1.

FODOR, Jerry A. and LEPORE, Ernest. 1992. *Holism: A Shopper's Guide*. Oxford: Basil Blackwell Publishers.

—. 1996. “All at Sea in Semantic Space: Churchland on Meaning Similarity.” *The Journal of Philosophy* 93(8): 381–403. Reprinted in Fodor and LePore (2002, 174–199), doi:10.2307/2564628.

—. 2002. *The Compositionality Papers*. Oxford: Oxford University Press, doi:10.1093/oso/9780199252152.001.0001.

FRANCKEN, Jolien C. and SLORS, Marc. 2014. “From Commonsense to Science and Back: The Use of Cognitive Concepts in Neuroscience.” *Consciousness and Cognition* 29: 248–258, doi:10.1016/j.concog.2014.08.019.

GENON, Sarah, REID, Andrew, LANGNER, Robert, AMUNTS, Katrin and EICKHOFF, Simon B. 2018. “How to Characterize the Function of a Brain Region.” *Trends in Cognitive Science* 22(4): 350–364, doi:10.1016/j.tics.2018.01.010.

JANSSEN, Annelli, KLEIN, Colin and SLORS, Marc. 2017. “What is a Cognitive Ontology, Anyway?” *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 20(2): 123–128, doi:10.1080/13869795.2017.1312496.

KANT, Immanuel. 1998. *Critique of Pure Reason*. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press. Translated and edited by Paul Guyer and Allen W. Wood, doi:10.1017/cbo9780511804649.

—. 1999. *Correspondence*. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press. Edited by Arnulf Zweig.

—. 2007. *Anthropology, History, and Education*. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press. Edited by Günter Zöller and Robert B. Louden, doi:10.1017/CBO9780511791925.

LEE, Jonathan L. C., NADER, Karim and SCHILLER, Daniela. 2017. “An Update on Memory Reconsolidation Updating.” *Trends in Cognitive Science* 21(7): 531–545, doi:10.1016/j.tics.2017.04.006.

MEEHL, Paul E. 1967. “Theory-Testing in Psychology and Physics: A Methodological Paradox .” *Philosophy of Science* 34(2): 103–115, doi:10.1086/288135.

NATHAN, Marco J. and DEL PINAL, Guillermo. 2017. “The Future of Cognitive Neuroscience? Reverse Inference in Focus.” *Philosophy Compass* 12(7), doi:10.1111/phc3.e12427.

OPEN SCIENCE COLLABORATION. 2015. “Estimating the Reproducibility of Psychological Science.” *Science* 349(6251), doi:10.1126/science.aac4716.

OVSIEW, Fred. 2014. “The Zeitraffer Phenomenon, Akinetopsia, and the Visual Perception of Speed of Motion: A Case Report.” *Neurocase* 20(3): 269–272, doi:10.1080/13554794.2013.770877.

PESSOA, Luiz, MEDINA, Loreta and DESFILIS, Ester. 2021. "Mental Categories and the Vertebrate Brain: the Neural Basis of Behavior." OSF preprint, doi:10.31219/osf.io/8cmhg.

PICCININI, Gualtiero and CRAVER, Carl F. 2011. "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches." *Synthese* 183(3): 283–311, doi:10.1007/s11229-011-9898-4.

POLDRACK, Russell A. 2006. "Can Cognitive Processes be Inferred from Neuroimaging Data? ." *Trends in Cognitive Science* 10(2): 59–63, doi:10.1016/j.tics.2005.12.004.

—. 2010. "Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed? ." *Perspectives on Psychological Science* 5(6): 753–761, doi:10.1177/1745691610388777.

PRICE, Cathy J. and FRISTON, Karl J. 2005. "Functional Ontologies for Cognition: The Systematic Definition of Structure and Function." *Cognitive Neuropsychology* 22(3): 262–275, doi:10.1080/02643290442000095.

RIZZO, Matthew, NAWROT, Mark and ZIHL, Josef. 1995. "Motion and Shape Perception in Cerebral Akinetopsia." *Brain: A Journal of Neurology* 118(5): 1105–1127, doi:10.1093/brain/118.5.1105.

SCHAEFER, Alexander, KONG, Ru, GORDON, Evan M., LAUMANN, Timothy O., ZUO, Xi-Nian, HOLMES, Avram J., EICKHOFF, Simon B. and YEO, B. T. Thomas. 2018. "Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI." *Cerebral Cortex* 28(9): 3095–3114, doi:10.1093/cercor/bhx179.

TAMAR, Diana I., THORNTON, Mark A., CONTRERAS, Juan Manuel and MITCHELL, Jason P. 2016. "Neural Evidence that Three Dimensions Organize Mental State Representation: Rationality, Social Impact, and Valence." *Proceedings of the National Academy of Sciences of the U.S.A.* 113(1): 194–199, doi:10.1073/pnas.1511905112.

UTTAL, William R. 2001. *The Limits of Phenology. The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, Massachusetts: The MIT Press.

YARKONI, Tal, POLDRACK, Russell A., NICHOLAS, Thomas E., VAN ESSEN, David C. and WAGER, Tor D. 2011. "Large-Scale Automated Synthesis of Human Functional Neuroimaging Data." *Nature Methods* 8(8): 665–670, doi:10.1038/nmeth.1635.

YEO, B. T. Thomas, KRIENEN, Fenna M., EICKHOFF, SImon B., YAAKUB, Siti N., FOX, Peter T., BUCKNER, Randy L., ASPLUND, Christopher L. and CHEE, Michael W. L. 2015. "Functional Specialization and Flexibility in Human Association Cortex." *Cerebral Cortex* 25(10): 3654–3672, doi:10.1093/cercor/bhu217.