University of Groningen

Modeling Mind and Matter

Kievit, Rogier A.; Romeijn, Jan-Willem; Waldorp, Lourens J.; Wicherts, Jelte M.; Scholte, H. Steven; Borsboom, Denny

Published in:
 Psychological Inquiry

DOI:
[10.1080/1047840X.2011.567962](https://doi.org/10.1080/1047840X.2011.567962)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kievit, R. A., Romeijn, J.-W., Waldorp, L. J., Wicherts, J. M., Scholte, H. S., & Borsboom, D. (2011). Modeling Mind and Matter: Reductionism and Psychological Measurement in Cognitive Neuroscience. *Psychological Inquiry*, 22(2), 139-157. [938333439]. <https://doi.org/10.1080/1047840X.2011.567962>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Modeling Mind and Matter: Reductionism and Psychological Measurement in Cognitive Neuroscience

Rogier A. Kievit , Jan-Willem Romeijn , Lourens J. Waldorp , Jelte M. Wicherts , H. Steven Scholte & Denny Borsboom

To cite this article: Rogier A. Kievit , Jan-Willem Romeijn , Lourens J. Waldorp , Jelte M. Wicherts , H. Steven Scholte & Denny Borsboom (2011) Modeling Mind and Matter: Reductionism and Psychological Measurement in Cognitive Neuroscience, *Psychological Inquiry*, 22:2, 139-157, DOI: [10.1080/1047840X.2011.567962](https://doi.org/10.1080/1047840X.2011.567962)

To link to this article: <https://doi.org/10.1080/1047840X.2011.567962>

Published online: 08 Jun 2011.

Submit your article to this journal [↗](#)

Article views: 535

View related articles [↗](#)

Citing articles: 3 View citing articles [↗](#)

REPLY

Modeling Mind and Matter: Reductionism and Psychological Measurement in Cognitive Neuroscience

Rogier A. Kievit

Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

Jan-Willem Romeijn

Department of Philosophy, Groningen University, Groningen, the Netherlands

Lourens J. Waldorp, Jelte M. Wicherts, H. Steven Scholte, and Denny Borsboom

Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

According to Karlin (1983), “the purpose of models is not to fit the data but to sharpen the questions” (Krukow, Nielsen, & Sassone, 2008, p. 3782). Given the rich and insightful commentaries we received, our approach to the reduction problem can be considered a success in this respect. The commenters have taken our ideas and expanded them both in breadth and depth. They have also critically examined the assumptions of our approach. In general, the commentaries suggest that the implementation of conceptually guided psychometric models is viable, is empirically tractable, and can be improved and revised on the basis of empirical and conceptual advances. Most important, they show that psychometric models yield increased depth and precision in dialogues concerning the foundational questions of cognitive neuroscience. In this rejoinder, we address the core points of criticism and present an expansion of the ideas we formulate in the Kievit et al. (this issue) target article, based on the ideas and suggestions offered by the commenters. Our focus is on the following set of themes that figured centrally in the comments: (a) What is the role of mechanisms with respect to our approach, (b) what explanatory levels should we study; (c) why should we engage in reductive science in the first place, (d) how can psychometric models be extended, (e) what interpretations of causality and realism are relevant for psychometric models, and (f) what philosophical positions can be translated into measurement models.

Mechanisms and Measurement Models

Perhaps most critical of our enterprise are Burnston, Sheredos, and Bechtel (this issue). They summa-

rize their criticism as follows: “Our central objection to the psychometric approach deployed by Kievit 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” (p. 108). We agree that mechanistic explanations are a worthwhile goal. This is as true for cognitive neuroscience as it is for other disciplines that deal with different explanatory levels (e.g. behavioral genetics, molecular chemistry). However, despite our concurrence with this overarching goal, we doubt whether the arguments of Burnston et al. (this issue) are relevant to our central point. First, the two endeavors—discovering mechanistic explanations and formalizing correct measurement models—are not mutually exclusive goals in science. In fact, they are mutually reinforcing. Hence, that mechanistic explanations are important does not imply that measurement models are unimportant, or vice versa. Second, the quality of mechanistic interpretations is essentially *dependent* on the quality of the measurement theoretical foundation used to relate psychological and neural phenomena. This is simply because the quality of *any* empirical study depends on the quality of its measurements.

We consider the second point first. As Bagozzi puts it, “Substantive researchers do not always develop the conceptual meaning underlying measurement relations, but they are implicit in the theoretical development of hypotheses and deserve explicit consideration” (p. 98). That is to say, the quality and validity of inferences about mechanisms *rely* on a valid measurement theoretical foundation, and by ignoring the measurement issues we may produce incorrect inferences about mechanisms. The dependency of mechanistic explanations upon correctly specified measurement

models also holds within psychology: When we find a consistent, replicable difference in performance across participants in a particular test (e.g., perceptual organization), we would ideally like to explain why this is the case. However, this question is secondary to, and reliant upon, the question of what the phenomenon (perceptual organization) is, how we may measure it reliably and validly, and whether our measurements allow for correct insights that form the basis of mechanistic explanations.

This dependency is especially relevant for neuroscience. For instance, consider the research discussed by Burnston et al. (this issue). Kanwisher, McDermott, and Chun (1997) examined the role of the Fusiform Face Area (FFA) in face perception. Both the initial claims (that this region of the brain was selectively active for stimuli of faces) and the ensuing refinements in other publications (that it may be better considered an area selectively active for stimuli for which a person can be considered an expert, e.g., perception of faces, or birds for bird-watching experts) of this research are important. However, this research clearly depends on a constellation of assumptions regarding the measurements on which these interpretations are based. In fact, throughout the history of cognitive neuroscience, a large body of research has been aimed specifically at examining the assumptions of empirical (neuroscientific) research, so as to ensure that inferences drawn from such studies are valid. For instance, studies have examined the relationship between neural activity and the Blood Oxygenated Level Dependent response (Logothetis, Pauls, Augath, Trinath, & Augelman, 2001), inter- and intraindividual variability in the Hemodynamic Response Function (the form of the temporal response of the increase in oxygenated blood; e.g. Handwerker, Ollinger, & D'Esposito, 2004), the functional and structural heterogeneity across individuals (Poline, Thirion, Roche, & Meriaux, 2010), inferences about individuals based on repeated measurements on similar trials (Fischer et al., 2003), and statistical inferences of neuroimaging data (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). For an excellent overview of these and other such issues, see Schleim and Rosier (2009).

As the aforementioned shows, these issues are well known and well documented in the field of cognitive neuroscience, and practicing researchers (such as Kanwisher et al., 1997) are generally well aware of their importance. For instance, if we were to (incorrectly) assume that the Hemodynamic Response Function is uniform across and within individuals, it may be that we discover spurious (de)activation in certain regions that is actually an artifact of an incorrect statistical model, and then offered a (incorrect) mechanistic explanation of the processes that are the source of this differential activity. The point is not that these assumptions invalidate research such as the study by Kanwisher et al.;

the point is that any mechanistic interpretation relies upon a constellation of such assumptions.

For instance, consider the (mechanistic) explanation of sex differences in perceptual organization (PO) by appealing to a sex-linked gene (e.g., Bock & Kolakowski, 1973; Posthuma et al., 2003). Being able to demonstrate that a sex gene can indeed explain sex differences in PO depends crucially on the measurement of PO (that is to say, does the PO test display measurement invariance? Meredith, 1993). Specifically, one has to demonstrate that the sex differences in the responses to the items measuring PO are attributable solely to differences with respect to PO, a latent variable. This is a basic measurement problem: Is the test measurement invariant (or unbiased) with respect to sex? If this is not the case, the observed sex difference on the PO test cannot be related unambiguously to the latent variable PO (i.e., there may not "actually" be a sex difference in PO ability), and we may end up trying to construct a mechanistic explanation for a difference that is not, in a meaningful sense, "there." If we cannot relate the observed sex difference unambiguously to sex, then we cannot know whether a demonstrated effect of the sex-linked gene is actually explaining a sex difference in latent variable PO. This example demonstrates the dependence of explanation on measurement.

The same issue may hold for the integration of neuroscientific and behavioral measurements. Although some well-established, mechanistically interpreted neuroscientific findings are rooted in large bodies of empirical data, it may well be that they require revision or refinement if the relationship between measurements and inferred traits is different for other subpopulations. For instance, there is evidence to suggest that neuroscientific findings such as the neural response to object processing and self-related thoughts are substantially different across cultural populations (Chiao & Cheon, 2010). Despite this fact, there have been few studies that explicitly compare neuroscientific measurement models for measurement invariance, which means that at least some mechanistic interpretations may be incorrect (Chiao & Cheon, 2010; "No research to date has explored comparisons in neural functioning of individuals living in small-scale versus industrialized societies," p. 89).

Finally, Burnston et al. rightly praise the advances and refinements of our insight into the function of the FFA after the initial findings by Kanwisher et al. (1997). However, the refinements of the original FFA theory based upon subsequent research (indeed an excellent example of heuristic identity theory [HIT]; McCauley & Bechtel, 2001) can also be tackled by means of correctly specified psychometric models. For instance, the subsequent research (following up on the study by Kanwisher et al.) that Burnston et al. discuss showed that, contrary to the initial hypothesis, the FFA

Figure 1. Using structural equation modeling to examine the presumed function of the Fusiform Face Area (FFA). *Note.* Here we have two latent traits: whether an individual perceives a face (measured by P- and N-indicators), and whether an individual perceives a bird (assuming expertise in bird perception, after the examples discussed in Burnston, Sheredos, & Bechtel, this issue). If activity in the FFA is truly only selective for face-stimuli, then activation in the FFA should be determined by the latent trait “bird perception.” If we specified this model, then a significant crossloading between the FFA and both latent traits would lead us to revise the original hypothesis concerning the FFA.

also shows differential activity to nonface stimuli (e.g., to other visual stimuli in which a person can be an expert, such as bird perception). In Figure 1, we show that this fact could have been detected by means of a correctly specified psychometric model.

The refinement of the theory could have been detected by showing there exist significant crossloadings of the FFA to two distinct latent variables (e.g., face perception and bird perception). That is, if the FFA is not simply a measurement of face perception, we expect the crossloading λ_{21} to be significantly different from 0. To the extent that the simple explanation of the FFA can and has been revised by further research, it can (or could have been) revised by better measurement models also: Covariation of the FFA with several distinct perceptual categories would lead to crossloadings of the FFA, revising the simple identity theoretical model, and thus lead to a more correct specification of the relationship between the unobserved latent trait and the observable measurements. This is an example how a correctly specified measurement model may be directly beneficial for mechanistic explanations.

Another way measurement models can inform mechanistic insights is because of the restrictive nature of the identity models. Neurological indicators that function as a reflective measurement of a psychological trait (or have substantial factor loadings in a MIMIC model) are worthy of closer examination, as such unidimensional covariation with the construct suggests the indicator may be in some way related to mechanisms relevant for the psychological process or trait. In this manner, a psychometric property of an indicator can inform and guide the search for (better) mechanistic explanations. The converse of course is also possible: Mechanistic insight, or mechanistic hypotheses, may

be invaluable in suggesting and selecting what indicators may be the best candidates for a measurement model of a given trait.

To summarize: We agree with Burnston et al. that mechanistic explanations are an important goal in cognitive neuroscience. However, even research aimed explicitly at uncovering mechanisms should be aware of measurement theoretical assumptions underlying the data that are subsequently interpreted as evidence for a given mechanism. Attempting to infer underlying mechanisms based on incorrect or unclearly specified measurement models may yield incorrect conclusions. Using the metaphor proposed by Burnston et al. (this issue): If we consider the data to be an epistemic inroad into the phenomena, we should make sure this road is well-lit.

Up and Down the Reductive Ladder

One of the core problems in cognitive neuroscience is the question at what lower level of description or granularity we should study people with respect to psychological phenomena. How do we decide whether to study coarse or detailed brain structure, neural firing rates, or protein synthesis?¹ As Vul (this issue) and Barrett (this issue) point out, the neurological measures we analyzed in our empirical illustrations (Kievit et al.,

¹ Although both Burnston et al. (this issue) and Vul (this issue) point out that one may also consider computational and algorithmic levels as different reductionist levels, for now we focus on the questions on, roughly speaking, the granularity or types of brain properties we should study, and why (although our approach is compatible with the algorithmic or computational level, more on the flexibility of N-indicators in a later section).

this issue) represent an arbitrary, and perhaps suboptimal, point to “stop”, and argue that those measures are too coarse to be plausible candidates for reduction.² Interestingly, arguing that the level we selected is either wrong or suboptimal implies that there may also be a “right” or at least a better level (for an excellent discussion on how to consider the nature of different levels, see Bechtel, 2008, p. 145.) The real question then becomes clear: How can we judge which level is the “right” one? By what criteria should we judge what indicators to use when engaging in empirical research? In order to judge what the “right” level is, we first need to discuss commenters’ answers to an even more fundamental question: Why engage in reductive science anyway?

Vul (this issue) discusses a possible way to judge the success of reductive science in connecting different explanatory levels: (a) A main goal in cognitive neuroscience is prediction (e.g., the predictive success of the accuracy of a surgery induced lesion), and (b) *given* this prediction criterion, we judge the success of a specific indicator (e.g., gray matter density) by its usefulness, or practicality, with respect to its predictive value. The reason that we shouldn’t try to measure, say, individual neurons within people is that this extra effort (assuming it is at all possible) will not increase predictive ability, at least not to the extent that justifies the extra effort. From the viewpoint of prediction, this is a coherent justification of reductive science; however, it is also quite limiting. It may still be worth engaging in reductive science if this is *not* the case, that is, when brain measurements do not outperform behavioral measurements of a given psychological construct.

For instance, Burnston, Sheredos, and Bechtel (this issue) argue that reductive science offers a way to get a grip on the (generating) mechanisms of the data. From this perspective, the criterion to judge the success or adequacy of a particular N-indicator is not necessarily that it outperforms behavioral measurements, but the extent to which it reveals or inspires hypotheses concerning the generative mechanisms that influence the higher level phenomenon (e.g., individual differences in general cognitive ability). From this perspective, the study of a particular brain property may be quite impractical or complicated, but nonetheless justifiable, as it yields “new” knowledge (into the generating mechanisms of the phenomenon of interest), not just predictive accuracy per se.

Barrett (this issue) takes a broader perspective on the purpose of reductive science, arguing that the pur-

pose of a neuroscientist is to “discover facts about the mind” by studying the brain, and that the “goals and proclivities of scientists may differ” (p. 121), a position compatible with research offering new kinds of insight without necessarily being more predictive or practical. Bagozzi (this issue), Vul (this issue), and Berkman and Lieberman (this issue) offer another possible goal for reductive science, namely, to cross-validate scientific constructs at other explanatory levels. If, say, a given psychological theory or construct (e.g., emotional responses to social exclusion) accurately predicts phenomena at other explanatory levels (e.g., changes in hormone levels following social exclusion), this may be considered corroborating evidence for the usefulness of such a construct at the psychological level. In this way, reductive science can offer a new kind of evidence for the justification of constructs that extends beyond the confines within which the psychological construct was developed.

A final, and pragmatic, illustration of the use of reductive science is offered by both Bagozzi (this issue) and Berkman and Lieberman (this issue), who argue that modeling of reductive science can be used to quantify and possibly overcome method bias in a new way. This is an excellent suggestion that we return to later.

In addition to this discussion on the specific utility of cognitive neuroscience, Haig (this issue) provides a meta-scientific structure for the goals of reductive science by embedding the issue in a larger framework, namely, the philosophical justification of science itself and the values by which we can judge the adequacy of scientific explanations. Haig discusses several approaches to judging superempirical virtues in science—superempirical because they go beyond the empirical adequacy of a specific line of research—for example, “unifying power” and “parsimony.” Which virtue one holds to be most important determines how one interprets the success of a given scientific endeavor. For instance, if the implicit value by which we judge scientific explanations is parsimony, then we may judge if, by this criterion, a neuroscientific description of a psychological trait succeeds.³ However, a practicing neurosurgeon collaborating with a neuroscientist may not care about parsimony: Whatever (measurable) constellation of neurological information concerning a patient best predicts where and how to conduct invasive surgery (which may be considered part of the virtue of fertility; C.F. McMullin, 1983) is adequate, no matter how much it lacks in elegance or simplicity. Similarly, if the superempirical value is the unifying power of a theory, then an overarching theory may lack in

²Although we agree that there are much more sophisticated and predictively better neurological indicators with respect to intelligence, we disagree with Vul’s assertion that intelligence is “not well predicted by coarse neural measures” (p. 139): Despite the coarse granularity, these measures together explain 25.1% of the variance in *g*, which is the higher end of explained variance in most research papers on similar topics.

³Haig (this issue) specifically addresses the importance of parsimony in comparing structural equation models. Although quantitative assessments are not the only interpretation of parsimony, the fit indices we use, (the AIC, the BIC, and the RMSEA) all take into account parsimony in some different ways (Claeskens & Hjort, 2008).

parsimony, but be justifiable nonetheless, assuming it succeeds in connecting a wide range of phenomena. For instance, despite the relatively simple basic principles, one would be hard-pressed to call the complete scientific corpus of the modern evolutionary synthesis (or the “extended evolutionary synthesis”; cf. Pigliucci, 2007) parsimonious. However, the unifying power of this theory, in offering predictions in fields as far apart as molecular genetics and behavioral ecology, is as large as one could wish from a scientific theory (of course the aforementioned depends on what notion of parsimony one wishes to adopt, a detailed discussion of which would lead us too far astray). It is clear from these examples that both the superempirical values by which one wishes to judge science in general as well as the perspective from which one engages in reductive (cognitive neuro-) science determines the framework from which one judges the adequacy of a particular line of research. There is no single correct perspective, and this fact alone implies that the assumptions of reductive (neuro)science should be made explicit in research to ensure that its relative success is judged on the basis of the correct merits.

Brains, People, and Reductionism

It is not uncommon in both the empirical and philosophical literature to find the following implicit argument: “Every psychological process occurs in the brain, the brain is material, therefore we should study the brain to understand psychology.” For instance, Barrett states, “In a certain sense, this has to be correct—unless you are a dualist, psychological measurements, *in the end*, have to be *causally reduced* to the brain” (p. 117). Strictly physically speaking, this is of course true. However, in terms of *explanation*, this need not necessarily be the case. Kim makes a similar point concerning physical reduction:

The core of contemporary physicalism is the idea that all things that exist in this world are bits of matter and structures aggregated out of bits of matter, all behaving in accordance with laws of physics, and that any phenomenon of the world can be physically explained if it can be explained at all. (Kim, 2005, pp. 149–150, as cited by Bagozzi, this issue, p. 88)

That said, one can take this discussion one step further and ask, Why stop at the level of the brain in the first place?

With respect to the explanation of psychological phenomena and the brain, there are two routes one may take with this argument: the first, which is often called strong reductionism, or physicalism, and the second, which we call pragmatic reductionism. Strong reductionism offers no *a priori* justification to stop at the explanatory level of the brain, as the brain is merely an

arbitrary stop in the hierarchical scheme of explanatory levels. Pragmatic reductionism can claim to stop at the level of the brain for pragmatic reasons, for example, because it offers predictive, intelligible, or insightful data, or because it has uncovered the generating mechanisms of the data.

For instance, if one argues in the tradition of (strong) physicalism like Kim, and argues that because people (displaying psychological phenomena) are physical, any (psychological) explanation must be explainable in physical terms, there is no reason to stop at the explanatory level of the brain. The problem becomes clear if we paraphrase Barrett’s comment as follows: “Brain measurements, in the end, have to be *causally reduced* to the protein structures (molecules/atoms ad infinitum)” (Bechtel, 2008, makes a similar point on p. 129).

Strong physicalism then, if interpreted in this manner, implies that *because* humans are material beings, we must study them at the lowest possible level to understand psychological phenomena. We propose that few practicing scientists would be comfortable with this position. To be sure, we do not assign this view to Barrett, we merely want to emphasize that the justification to stop at the level of the brain to understand or explain psychological phenomena needs to be made on the basis of other arguments than merely the fact that people are physical objects. We propose that few practicing scientists would be comfortable with this position. Even eliminativists such as Churchland, as discussed by Bagozzi, tend to argue that psychological phenomena are not “real” but are “actually” neural or hormonal properties of the brain. Why these properties in turn then should not be eliminated, or are taken to be more real, is rarely spelled out more explicitly than a passing reference to future developments.

The second line of reasoning seems much more viable: the pragmatist reductionism account. This account offers us grounds to stop at the explanatory level of the brain (e.g., gray matter density, cortical activity) because it “works,” that is, it offers predictive, sometimes explanatory, and insightful perspectives concerning the higher order property. The pragmatist reductionism account implies that any claim that we *should* study the brain is fine, as long as it is grounded in empirical justification, not in *a priori* claims concerning the “materialness” of people. One cannot, in our view, use a principled argument to study lower explanatory levels and then “get out” halfway. Take, for instance, Churchland (1981): “Our mutual understanding and even our introspection may then be reconstituted within the conceptual framework of *completed neuroscience* [*italics added*]” (p. 67). If one wishes to be an eliminativist about folk psychological constructs, then it seems inconsistent to rest our hopes upon a “completed neuroscience”: One must also be an eliminativist about neuroscientific explanations with respect to molecular explanations, atomic explanations, ad infinitum.

The focus of Churchland on *neuroscientific* reduction then is tenable if he can support claims concerning the imminent demise of folk psychological explanations to neurological properties by empirical evidence. These may well be forthcoming, but they cannot be claimed on a priori grounds. A possible source of such arguments is an inherent bias to favor lower, biological explanations over “higher” explanations. Kendler (2005) addresses this type of implicit preference with respect to genes, the environment and psychopathology:

I suggest that we feel comfortable with “X is a gene for Y” and not “A is an environment for B” because we implicitly assume that genes have a privileged causal relationship with the phenotype not shared by environmental factors. However, empirical evidence does not support the position that genes code specifically for psychiatric illness while the environment reflects nonspecific “background effects.” (p. 1248)

The bias for the lower order properties seems to stem more from intuition than empirical fact. Although psychologists have long accepted error in their measurements as unavoidable, some discussions seem to assume that the errorless measurement of neural processes is less problematic. We disagree with this notion, and find the idea that neural processes (especially the type of neuroscientific abstractions that figure in explanatory accounts) can be directly measured without error equally problematic as the idea that psychological phenomena can be measured without error (this largely resolves the difference in representation between us and Barrett (this issue): We, for the time being, consider the two measures on equal footing precisely because this errorless measurement of the relevant neural property is not forthcoming).

In our view, the route of pragmatic reduction is most useful: The extent to which certain properties can be reduced or explained by lower levels is an open question that should be adjudicated based on empirical evidence and correct, intelligible, predictive explanations. It may be that constructs such as “depth perception” can completely be reduced (either as laws or as fully developed mechanistic explanations) to neuroscientific explanations, whereas others (intelligence, working memory capacity) will resist such wholesale reduction. In our view, empirical and conceptual advances (such as HIT or our approach) are more likely to be of use than a priori arguments concerning the likely end result.

Extending Psychometric Models

The commenters offer a host of suggestions to improve and extend the implementation of our models. Bagozzi and Barrett in particular take our SEM rep-

resentations and expand and adapt them in various insightful and useful ways. Before we discuss possible alternative models or extensions, it is important to clarify two misunderstandings concerning the nature of the N-indicators, namely, which N-indicators can be used in psychometric models, and the status (data/phenomenon) of complex correlations among indicators. Vul (this issue); Burnston, Sheredos, and Lieberman (this issue); and Barrett (this issue) argue, based on the indicators we used in our analysis, that it is unlikely that activity in a single brain region, or gray matter density in a particular area, will ever be identical to a complex trait such as intelligence and is therefore unlikely to display an identity theoretical relationship (i.e., for a reflective model to fit). The commenters suggest that, for this reason, the identity theoretical model in our approach was in a sense “doomed,” given the complexity of intelligence versus the relatively coarse nature of our neural indicators, and that this simplified assumption (brain area = cognitive faculty) may jeopardize the enterprise of modeling in this way. We fully agree with the position that it is unlikely that single brain areas will map one-to-one onto cognitive faculties. In fact, our approach can be seen as an attempt to circumvent this (implicit) type of thinking. The fact that several authors raise this criticism means we have been unclear about the nature of the N-indicators, and we wish to remedy that here.

The N-indicators in structural equation models can represent *any* neural (or physiological) measurement, regardless of whether they are direct or indirect, coarse or fine-grained (Bechtel & Mundale, 1999). They may include, but are not limited to, a diverse range of indicators such as cortical thickness, network measures of “functional connectivity” of a particular brain area, a Fourier power transform of an EEG measurement of a given source, the proportion of gray matter compared to white matter in the brain of an individual, network coherence in resting state measurements, neurotransmitter levels, or any of the wide range of other measurements commonly acquired in cognitive neuroscience and related disciplines. The flexibility of the N-indicators is, in our view, one of the strengths of psychometric models. It is precisely the shift of focus from what, or where, the neurological measurement “are” to what they “do” in a measurements model that makes the measurement theoretical questions explicit, and may steer us clear from the dangers of the neophrenological approach everyone wishes to avoid.

For instance, recently there have been several recent developments in mapping the small world characteristics of networks in the brain and their relationship to psychological phenomena (Bassett & Bullmore, 2006; van den Heuvel, Stam, Kahn, & Hulshoff Pol, 2009). Small world networks (Watts & Strogatz, 1998) are usually defined by two parameters: small average path

length and a preponderance of clusters (i.e., a high clustering coefficient). Although these parameters are indirect and abstract measurements of brain properties, they can figure as N-indicators in a structural equation model. For instance, van den Heuvel et al. (2009) showed that there appears to be a correlation between average path length and intelligence, that is, people with shorter average path length generally had higher IQ scores. In our approach, one may take these parameters and use them as N-indicators in a reflective model. This allows one to see whether these indicators can figure as reflective measurements of intelligence, that is, whether the small world characteristics are a function of the same property as the IQ scores. It is precisely because psychometric models can naturally accommodate such diverging measurements that they can be so useful: It should be noted that the general latent variable model is the *only* model in existence that can empirically specify the hypothesis that two distinct variables measure the same property. Thus, what matters is how indicators “perform” in a measurement model, what properties they share, and what variance they explain or predict. If indicators of neural properties are able to perform these tasks, they can figure in models.

A second issue concerning our indicators need also be addressed, namely, the suggestion that we are merely interested in observable indicators. Burnston, Sheredos, and Bechtel (this issue) argue that instead of explaining phenomena, we are only explaining data (c.f. Bogen & Woodward, 1988). We disagree. The data are the observations as coded in a data file. To explain their existence, it suffices to read the Materials and Procedures section in a paper. Here we are explaining not the data but the structure of a covariance matrix of psychological and neuroscientific measures. The structure of such a covariance matrix is a complex empirical phenomenon if anything is. Now, in setting up a model, we attempt to explain the structure of the covariance matrix by postulating specific hypotheses on sources of variation in the measures (in the words of Barrett [this issue], “all measurement questions are also philosophical questions about how variation in numbers hint at or point to reality,” p. 115). In doing this, we aim to map the sources of variation to the empirical patterning of the covariance matrix. Surely this goes beyond the directly observable data, if only because it requires one to postulate and evaluate the quality of measurement theories (see also Haig, 2005). It may not be possible to clearly demarcate the domains of data and phenomena very sharply—in fact, such a delineation may be illusory anyway—but in our account we certainly show awareness of the fact that phenomena are different from the data. In addition, establishing measurement properties of utilized instruments is clearly a necessary step toward understanding and explaining phenomena. This process in-

volves the specification of accurate measurement models, which in turn requires hypotheses on the processes generating the data. The history of the development of psychometric models has shown that this iterative process is mutually beneficial in the best tradition of HIT (McCauley & Bechtel, 2001).

Bagozzi and Barrett offer several substantial extensions of our models. They take our ideas and show how different SEM implementations can be used to tackle complex empirical phenomena and other philosophical theories. Bagozzi takes our ideas and generalizes our approach even further. Where our approach aims to translate existing philosophy of mind theories into models that represent the conceptual implications of those theories, Bagozzi proposes a general, or skeleton, model (Bagozzi, this issue, Figure 5) that can be modified to represent a variety of conceptual hypotheses. This could be considered a meta-psychometric approach to philosophy of mind. From this perspective, the mental state (M1) and the physical state (P1) of a person are latent variables, and both measure the theoretical property (T1) of that person. The advantage of this overarching model is that this model can be adapted to represent and test a variety of conceptual hypotheses. Various conceptual assumptions can be implemented by means of standard psychometric constraints, such as setting parameters to certain values, equating latent variables, and constraining allowable correlations. This is a versatile extension of our original approach, and an insightful manner to underscore the notion that for many philosophical theories concerning the relationship between psychological and biological properties it is at least in principle possible to translate them into empirical theses. Of course, there may still be disagreement on how to best represent a particular theory (e.g., Bagozzi and Barrett both propose alternative representations of supervenience theory), but in our view, it is exactly this type of disagreement that may yield maximal payoff, both philosophically and empirically.

Bagozzi also argues that neurological measurements can be implemented in a latent variable model “by themselves.” That is to say, a unidimensional latent variable can be formed from covarying neurological measurements, which if it correlates 1 with a psychological variable, can be considered as evidence for identity theory.⁴ In fact, modeling neurological indicators “by themselves” by means of an exploratory factor analysis may do more justice to our current prevailing ignorance concerning the structure of

⁴Although the suggestion by Bagozzi to test the deviation of the correlation from zero instead of the full reflective model is interesting (and his comments concerning method factors valid), we prefer a model with a single factor. By representing the latent construct as one latent variable, we also imply certain interpretations in terms of possible interventions that are different from having two separate, albeit highly correlated, latent factors.

neurological covariation. This is an excellent suggestion, and an approach we have followed and implemented in more recent work (Kievit et al., 2011), where we examine intelligence and a more complex constellation of brain measurements than discussed here. In this study, we take an exploratory factor analytic approach with respect to the N-indicators, to uncover the structure of these properties, and correlate them with the psychological constructs of interest. Using an exploratory factor analytic approach for different constellations of neurological measurements in future research may provide useful insight into the structure of interindividual differences in neural makeup.

Barrett discusses a wide range of new models, largely from the perspective of psychological construction (Barrett, 2009, this issue; Coan, 2010), a position naturally compatible with formative models. In this and previous work, Barrett shows how the biological study of emotion research has recently developed several quite sophisticated representations of the relationship between the physiological components of emotions and the psychological, constructionist positions. Barrett illustrates how complex empirical phenomena, such as emotions, can be tackled in this manner and offers both simple and hierarchical models to support this view. This offers the possibility of reconciling both the distinct nature of psychological and neurological properties while attempting to bridge the divide in an intelligible manner. It will be interesting to see both how the constructionist position fares in emotion research, and to what extent the ideas will extend into other fields of psychological inquiry.

The Importance of Behavior

Berkman and Lieberman (this issue) argue that we should include measurements of behavior in our models, significantly extending the scope of our approach. Along with Bagozzi, they argue that behavioral indicators should be included in our models, as they represent important (and often neglected) aspects that are relevant for psychological explanation. Berkman and Lieberman take the supervenience model and the associated characteristic of multiple realizability theory and use it to explain the relationship between observable behavior and psychological characteristics. The importance of behavior in such models, and the role of multiple determination, is also suggested by Barrett and Bagozzi. Barrett (2006; this issue), for instance, discusses several examples where the one type of overt behavior can be caused by different constellations of (presumed) psychological states. For instance, in animal behavior research, certain behaviors (e.g., freezing behavior) may be the result of quite different constellations of psychological states (avoiding being seen, paralyzed by fear, or the moment of readiness before

Figure 2. Supervenience “deluxe.” An extended supervenience model. *Note.* P indicators are only drawn from the first psychological variable for visual clarity. This model represents hierarchical multiple realizability. The same behavioral state (e.g., “freezing behavior”) can be determined by different psychological latent trait values, which in turn can be determined by different constellations of neural activity and/or constitution.

a fight-or-flight response). Although the model proposed by Berkman and Lieberman is not statistically identified, it is illuminating to think about the inclusion of other explanatory levels in such ways. An alternative model based on Berkman and Lieberman (this issue), presented in Figure 2, could be implemented to examine the predictions concerning determination and multiple realizability of behavior with respect to psychological indicators and neurological indicators in the same model. The constellation of psychological components that lead to a given behavior is generally quite complex. If only the behavioral outcomes are studied, we may not be aware that different constellations of psychological primitives can lead to the same type of behavior. Similarly, by only studying psychological processes or dispositions, we may be unaware of the constructive role they play in behavior. Modeling several explanatory levels simultaneously may yield benefits for several fields of study. This interpretation may be referred to as a hierarchical supervenience model, and is similar to the constructionist model Barrett proposes in Figure 7B.

Models as the aforementioned may provide structure to connect explanatory levels in an empirically tractable manner. For instance, reductive science generally crosses only one “divide”: Brains are informative about individuals but rarely about large social groups. Similarly, the mapping of specific genes (and gene expressions) has been relatively successful for lower order phenotypic properties (e.g., serotonin transporter gene; Heils et al., 1996) but rarely for higher order

properties (e.g. schizophrenia and genetic loci; Penke, Denissen, & Miller, 2007). By linking different explanatory levels and studying the explained variance at each level through hierarchical models, we may be able to derive the predictive abilities across relatively large explanatory gaps (e.g., groups to brains, behavior to genes) by studying the explained variance at each level. In this way, it could actually be possible to connect different explanatory levels. Linking these different explanatory levels by means of measurement models is a tractable empirical approach that can serve to link different manners of inquiry.

Mismatch of Latent Variables: Sources and Solutions

Berkman and Lieberman (this issue) argue for the inclusion of behavior into models for a variety of reasons. One of these reasons is the common mismatch between indicators of constructs that seem similar, when estimated by means of different indicators. They argue convincingly that, often, psychological predictors (i.e., self-reporting of likely behavior) do not match observable or measurable behavior (objectively observed ratings of behavior). The issue of measurements of the same construct not matching is a more general problem that also applies when comparing neural and psychological levels (consider, e.g., Bagozzi's proposal to model a latent variable with only neural indicators and a latent variable with only psychological indicators, and the correlation deviating from 1). Given our focus on integrating different explanatory levels, such mismatches are worthy of close attention.

There are a variety of reasons for a mismatch between latent factors. We discuss four reasons that are especially relevant for our approach and can be addressed using psychometric models. First, the lack of correspondence may simply be due to noise, that is, poor measurement of the one "true" latent trait of interest. In other words, a given score on a self-report test concerning the likelihood of "going to parties" or "engaging in conversation with strangers" may be a quantitatively worse indicator of the same latent trait, Extraversion (e.g., McCrae & Costa, 2004), than behavioral measurements, as it introduces additional noise. If this variation is random around a given "true" mean (i.e., normally distributed error), this is not necessarily a problem: We can still model the same latent trait with two classes of indicators, but the psychological measurements will have lower factor loadings (and higher error terms). In such cases, one may still consider the response to the self-report question, "How likely are you to engage in conversation with strangers" as a *measurement* of the latent trait Extraversion, albeit not so good a measurement as actually measuring behavior of that person in a social situation.

The second explanation for mismatch is that the factors measured by means of self-report and those measured by "actual" behavior both represent coherent latent factors, albeit of different constructs. For instance, a person may have a factor score on the self-report based questionnaire measuring extraversion (Extraversion self-report, or ESR) and a factor score based on a measurement tool that measures behavior in a set of relevant social situations (Extraversion behavioral report, or EBR). If these are not perfectly correlated (i.e., the deviation of the correlation from 1 is significant), we could consider these measurements of two distinct latent traits, namely, "self-perceived extraversion" and "extraversion-behavior," both of which may be psychologically relevant (the first with respect to theories of self-perception, the second for dominance theories, or group behavior theories). Of course, the lack of a perfect correlation does not in itself prove that we should consider both the latent variables to be sensible constructs; however, this can be the case. If we have reason to believe or argue that there is only one "real" trait of extraversion, we should carefully consider and attempt to quantify the relative method bias of measuring this trait behaviorally and psychologically. However, if they actually represent two relevant psychological dimensions, then treating the different scores on these latent variables as the consequence of method variance is a mistake. The extent to which the two constructs are both allowed to figure in a psychological ontology may depend on a host of other factors such as intelligibility, correlations with relevant other characteristics, predictive ability, and cross-validation.

A third possible source of a lack of correspondence between P and B indicators (or factors) is the influence of the environment, as Berkman and Lieberman (this issue) suggest. The influence of the environment can be modeled and made explicit in a psychometric model. One may consider a case where a researcher wants to manipulate an environmental variable such as "high pressure" or "low pressure" (e.g., giving a talk for 100 strangers or one friend) to see how it influences "extraversion" of participants in a social setting (i.e., the EBR score). Instead of considering this environmental influence as something that will necessarily create a mismatch between the ESR and EBR scores, we can attempt to model the influence of this environmental factor as an exogenous variable. For instance, we could construe "high pressure" versus "low pressure" as a dichotomous influence on the relation between self-reported extraversion and behaviorally observed extraversion in a certain setting. The influence of such environmental variables can easily be incorporated in a psychometric model, as is illustrated Figure 3.

If a dichotomous exogenous indicator has a certain influence on the EBR factor score, the nature of this relationship as estimated by means of a psychometric

Figure 3. Estimating environmental influence on a latent trait. *Note.* We may test the hypothesis that an external influence can affect how people behave in certain situations, and that if we do not take this variation into account, this may lower the correlation between self-reported extraversion (ESR) and behaviorally observed extraversion (EBR).

model can be insightful. For instance, a high-pressure situation may increase the variance of the factor (people differ more greatly in extraversion in high-pressure situations, decreasing the correlation between the two latent factors) or have an intercept effect on the factor mean (everyone becomes shyer in a high-pressure environment) or a host of other psychologically and behaviorally relevant possibilities. We can test this hypothesis with a multigroup factor model with a binary exogenous manipulation (high/low arousal as an experimental manipulation induced in two randomly selected groups, pertaining to the situation in which a person may or may not display outgoing behavior) that influences the latent behavioral trait. In this manner it is possible to quantify the influence of the environment on the behavioral factor and to partially explain the lower correlation between the two latent traits. Not only does this method increase insight, it also allows us to formalize explicit hypotheses concerning the status of traits and how we expect them to (co)vary, and to refine our insight into the sources of mismatch between psychological and behavioral properties. This again represents a case where psychometric models not only allow us to model certain plausible disturbances but also make the measurement theoretical assumptions of the researcher explicit.

A final explanation for we will discuss here a mismatch between latent variables is method bias, or method variance. Bagozzi proposes an additive trait-method-error model to model and formally quantify method bias. This is an excellent approach, as it allows both for cross-validation of constructs with different methods and formal quantification of method bias. This approach can be used as a way to decide certain arguments concerning the accuracy or quality

of measuring psychological constructs either in the traditional psychological manner or by developing a neuroscientific measurement model. For instance, Lamme (e.g., 2004) has argued that psychological measurement of consciousness or conscious awareness is inherently problematic, because of inherent bias in any psychological measurement that cannot be circumvented. In consciousness research, almost every experimental paradigm takes self-report as the gold standard. That is to say, specific parameters of stimuli, usually visual (e.g., clarity, presentation time) are manipulated. Participants are asked whether they perceive the stimulus, and when participants report “yes,” such trials are compared in terms of neurological measurements to trials where participants reported not being aware of a stimulus (implicitly assuming errorless measurement for the P indicators). However, it may be that asking people to report whether they were aware of a certain stimulus is a poor “measurement” of the phenomenon of interest, namely, whether the person was *actually* conscious of the stimulus at the time, because it takes a couple of hundred of milliseconds to report a stimulus. Only a limited amount of information is selected by attention, and in this way protected from loss. Verbal report, in such a scheme, is only related to the items saved by attention and does not necessarily address the full pallet of visual experience.

We propose a model to test whether self-report is biased with regard to consciousness of a stimulus in Figure 4. The purpose is to quantify whether biased self-report leads to correlated error terms, and so to a worse measurement of the latent trait. Here, the psychological indicators represent self-report or behavioral performance on a suitable task (such as error rate and reaction time on a change blindness paradigm).

Figure 4. Estimating method bias inherent in self-report. *Note.* Here we can test the hypothesis that the error terms of the psychological assessment of the latent trait are correlated. We can free the parameter and use a likelihood ratio test to test whether the fit of the model improves significantly.

The neurological indicators are those considered suitable or relevant for conscious awareness, in the case of Lamme usually the presence of recurrent processing from higher cortical areas to lower cortical areas (Lamme & Roelfsema, 2000). Here, the presumed psychological bias (in line with Lamme) is represented formally by the box “bias,” an exogenous influence that directly affects the correlation between psychological indicators of awareness.

This box of measurement bias is the cause of correlated residuals between psychological indicators, due to forgetting what was perceived, or post hoc rationalizations. For instance, people may want to give a “coherent” picture about what they saw, and answer as they think they “should” answer instead of what they actually saw. Or people may simply be “wrong” about what they believed they were conscious of several moments ago. If any of these problems occur in a nonrandom manner, they can adversely affect the quality of the measurement of the construct of interest, leading to correlated error terms and, thus, model misfit. Statistically, this can be tested by means of a model which frees the parameter that estimates the correlation between the error terms of P1 and P2, and seeing if it deviates significantly from 0 (conventionally, the correlations between error terms are set to 0). The problem of possible bias in psychological measurement of latent traits can then be tackled empirically with the model presented in Figure 4. The empirical quality of

indicators as they figure in such a model can then be quantified, and bias in either predictor can be estimated empirically (e.g., by examining correlated error terms of indicators). From this perspective we try to find N-indicators that function in this measurement model and that may ultimately outperform the P-indicators of consciousness (i.e., self-report), as N-indicators may not suffer from bias.

Realism and Naturalized Philosophy

Several commenters discussed our position concerning the reality or realist interpretation of the latent variables in our models. That is, to what extent do we consider latent variables to be “real,” or to constitute a feature of reality that is in some sense “out there,” independent of our models (e.g., Burnston et al., footnote 4). It would be unwisely ambitious to attempt to resolve the ontology of psychological attributes here in one fell swoop, and we do not attempt this within the confines of this reply. However, we can specify the kinds of relations in which psychological attributes should be able to enter for them to be plausible candidates for figuring in measurement models of the kind we proposed.

For the reflective model, the primary requirement is that differences in the attribute of interest act as a common cause (Reichenbach, 1956, chap. 19) of observed

differences in the test scores. For instance, differences in mercury, electrical, and digital thermometers across environments depend on a common cause, namely, differences in ambient temperatures. This means that the common latent variable in a measurement model for these thermometers can be justifiably interpreted in these terms. The reason is that the reaction of each of these instruments can be causally traced to the ambient temperature in the environment (this information is so prosaic that it can often be looked up in the user manual; a notable difference compared to the user manuals for psychological tests).

Any variable that is capable of playing the role of common cause can thus figure as a latent variable in some reflective measurement model (not necessarily a factor model). Such a situation has important ramifications. For instance, one can be *mistaken* about differences in the attribute (e.g., in the case of intelligence, it is possible that one concluded that John is more intelligent than Jane, whereas the reverse is in fact the case). Therefore, one can also be *right* about these differences. Clearly, whether one is right or not is *not* a function of the data (otherwise no measurement model would be needed) and this, in turn, means that the truth maker of a sentence like, “John is more intelligent than Jane,” must be an empirical feature of John and Jane’s standings on the attribute itself. In that sense, therefore, it is inherent to our use of the latent attribute that we assume the attribute exists: It must be able to function as a truth maker for hypotheses that bear on it. As MacCorquodale and Meehl (1948) say, such attributes “have a cognitive, factual reference in addition to the empirical data which constitute their support” (p. 106).

Is such a factual reference sufficient to call an attribute “real”? That is a question on which debate is possible. Consider Dennett (1991) on the question of whether “centers of gravity” exist:

Philosophers generally regard such ontological questions as admitting just two possible answers: either beliefs exist or they do not. There is no such state as quasi existence; there are no stable doctrines of semirealism (p. 27) and later, the question of whether abstract objects are real—the question of whether or not “one should be a realist about them”—can take two different paths, which we might call the metaphysical and the scientific. (p. 28)

Centers of gravity would seem to have pretty good credentials compared to most psychological constructs proposed today. They are measurable, they are determinate, one can be wrong about centers of gravity, and they can plausibly figure as common causes in accounts of distinct measurement instruments capable of detecting differences between them. Whether the ultimate philosophical verdict will allow them in the cabinet of actuality, or will demote them to the realm of useful fictions, is an open question on which the philosophi-

cal jury is still out (and is likely to remain so forever). However, they are “real enough,” so to speak, to figure in measurement models, and this is all we need to progress with empirically informative tests.

How often we can justify the assumption that psychological attributes are determinate in this sense? Can they function as common causes? These are open questions. We cannot resolve them, but we can make an interesting observation on why they are open. Namely, in many cases of psychological testing we do not have a clear idea of the processes that lead to test scores, and hence it is hard to trace back these processes to the level of the attributes measured. The introduction of psychological attributes as latent variables in reflective measurement models thus rests on something of a gambit, namely, that someday someone will be able to flesh out the processes that connect the observables to the attribute of interest. How to flesh out such process-level explanations is an active area of research (e.g., see van der Maas, Molenaar, Maris, Kievit, & Borsboom, in press). Interestingly, mechanistic explanations of item responses are likely to play a key role in this research.

Of course, observing a statistical pattern compatible with the existence of a psychological attribute in itself does not establish the reality of that attribute. However, insofar as a measurement model withstands empirical tests, the latent variables in that model are empirically supported. We may provisionally accept such attributes into our scientific ontology. Naturally, they can remain there only as long as the empirical predictions drawn from their existence are in line with empirical findings regarding the measurements, and as long as no better explanations of these findings are available. It is important to note that these predictions may also include predictions on what will happen after interventions. For instance, if we manipulate a latent variable, then we expect specific changes in its indicators (e.g., in a factor model these should be proportional to factor loadings). If we proposed a reflective latent variable of, say, working memory capacity, including psychological and neuroscientific indicators, a manipulation of the latent variable (working memory capacity) by some intervention should result in the predicted changes in *both* classes of indicators to be compatible with the realist interpretation of the latent variable.

Cross-validation at different levels of explanation, predictions outside of the framework where the construct was developed and a resistance to alternative explanations or descriptions may all bolster our provisional acceptance of a construct. The assumption that there is something real to be captured imperfectly by means of empirical inquiry seems to be both the most productive assumption, and most compatible with scientific inquiry in general (for a similar view, cf. Edwards & Bagozzi, 2000, p. 157). However, others do not necessarily have to agree with our perspective to be able to benefit from the possibilities offered by the

SEM framework. There is a wide range of perspectives on the reality of (latent) constructs within the psychometric, psychological, and philosophical communities, and researchers from different perspectives may still employ the same techniques, despite having diverging assumptions concerning the reality of constructs.

Causality and Structural Equation Models

Several commenters discuss the precise nature of the causal relationships we assume in our models. The central issue of this article, the relation between psychological and neurological data, is fundamentally a question about the implicit and explicit causal assumptions within reductive psychological research. However, it is important to remember that even scholars who agree on the fact *that* such models should be interpreted causally need not agree on the precise interpretation of the causal relations. In fact, causality is a rich and complex topic that has received much scholarly attention. Despite this attention, the literature has not yet converged on a universally accepted definition of causal relations, nor will we achieve that goal within this rejoinder.

For instance, some scholars have examined the concept of causality from a counterfactual perspective, which states that "A causes B" means as much as "If A had been different, B would have been different" (Lewis, 1973; Psillos, 2004). Others have emphasized mechanistic connections, in stating that "A causes B" means that there is a mechanism that connects A and B, in the sense that feeding A to the mechanism typically produces B as output (Machamer, Darden, & Craver, 2000). For a discussion on different statistical interpretations of mechanisms and causal intervention, see, for example, Holland (1986) and Rubin (1986). Still others have developed synergies of these two perspectives (Psillos, 2004; Woodward, 2002), or have developed technical demarcations, in terms of necessity and sufficiency, to explicate conditions required for common language interpretations of causality (Mackie, 1965). For an insightful discussion on causality with respect to measurement models, see Edwards and Bagozzi (2000). As is clear here, a wide range of causal perspectives are possible and can figure as the implicit foundation of empirical inquiry. However, we do not wish to uniquely fix the interpretation of causal relations; depending on the nature of a field of research either counterfactual, mechanistic, or interventionist interpretations may be most appropriate for a given measurement model.

The model that raises the most controversy is our interpretation of the MIMIC model, representing supervenience theory, as being a relation of causal determination. The precise interpretation of supervenience, as several commenters note, is the topic of a heated and

prolonged debate in the literature (e.g. Horgan, 1993; Kim, 1987; Lewis, 1994). Burnston et al. (this issue), for instance, argue that supervenience is a relationship of realization, not of causation, whereas Barrett (this issue) describes a causal interpretation of supervenience, drawing on Searle (1992). Even from the same author, Kim, a range of interpretations can be drawn. For this reason, as we note in our target article, we took the most straightforward, "old-fashioned" interpretation of supervenience. As Bagozzi notes, we adopt the terminology of supervenience as a relationship of determination and dependence from Kim. Burnston et al. also take issue with our interpretation of causality: "The relationship involved in supervenience is one of *realization*, not of causation." This implies a dichotomous "choice" that, in our view, need not be mutually exclusive. Just as the relationship between temperature in a room and the readings on a thermometer can be considered a relationship both of causal determination and of measurement, the same can be said of psychometric models.

Although we may disagree on the most sensible interpretation of the models, we are inclined to agree with the commenters that, regardless of the specific interpretation one chooses, the relation between the indicators and the latent variable in the formative part of the MIMIC model is different from the (causal) relationship between the latent variable and the reflective indicators. Our interpretation of the causal status of the formative indicators in a MIMIC model is perhaps best described by the term "mereological causation." That is, the lower indicators *determine* the higher level properties in a part-to-whole fashion, and if the lower properties had (counterfactually) been different, the higher property would also be different. This relation is a part-to-whole, counterfactual perspective on the relation between the component parts (neural indicators) and the higher order property. (See Bechtel, 2008, for an excellent discussion on the notion of levels and determination.) The lower level constitution determines the higher level phenomenon in the following sense: The lower order properties determine the higher order properties by virtue of forming its component parts, and if (components of) the lower order properties were different, or absent, the higher level property would, by necessity, also be different. If the structure of the brain of person X *would have been different* in these specific regions of interest, then this person would have also been different in this psychological construct of interest (assuming other component parts are held the same), and the extent to which this is the case follows from the parameter estimates. It is in this sense that we could consider, for instance, the constitution of the brain of a person to be causal of his or her position on some latent psychological dimension. It is important to note that within the framework of the MIMIC model, this work is generally not considered complete: The disturbance term ζ on the latent variable repre-

sents the totality of unmeasured causes not included in the current, incomplete version of the model. Diamantopoulos (2006) states, “Thus, the surplus meaning possessed by a formative construct relates to the influence of unmeasured causes, i.e. indicators not included in the model” (p. 14). That is, the disturbance term in the MIMIC model suggests that N-indicators (in our case) present in the model are considered a part of a larger domain of N-indicators that, if we were capable of finding them, together completely determine or cause the latent variable. It is in this sense that we can consider the ontological status of the latent variable in the MIMIC model as more than just the weighted linear summation of the N-indicators: In line with our previous section on realism, the latent variable represents a currently incomplete and imperfect representation of the true, or completed, latent variable, which we will probably not be able to completely map or attain.

It is clear that there are many sensible positions on the causal relationships in the various models, and we do not presume to favor any one of these interpretations in particular. The primary reason for this is that appropriate notions of causality and the causal interpretations of causal models vary across research contexts (as we discussed in our target article, inter- and intraindividual explanations differ quite radically). To summarize: In our view, the use of models such as the reflective and formative models previously discussed invites researchers to consider the question what causal interpretation may be sensibly attached to the “arrows” in the models. However, much like the issue of the realist status of latent variables, this does not require the researcher to commit to any one of these interpretations a priori in order to use such psychometric models.

Theoretical Extensions

In our target article, we focused on identity theory and supervenience but, as Bagozzi rightly notes, a range of other philosophical theories are relevant to our stance, including eliminativism, folk psychology, functionalism, and property dualism. It is encouraging to see the extent and depth to which Bagozzi and Barrett have adapted our line of reasoning to other philosophical positions, suggesting our approach can indeed be applied more generally. For this rejoinder we focus on two theories that are especially promising, namely, the distinction between type and token identity theory, and the theory of emergence.

Type/Token Identity Theory

Barrett draws attention to an important distinction, which we examine in more detail. Identity theory, as supervenience, comes in many flavors, compatible with

different measurement models. An important distinction is between type and token identity (e.g., Aydede, 2000; Fodor, 1974; Place, 1999; Rowlands, 1992). These two theses offer quite different perspectives on the nature of the identity theoretical relationship, and Barrett is correct in mentioning that our discussion focuses largely on type identity theory. The hallmark example of a type identical property is to equate “Being in pain” (X) with “C-fibers firing” (Y), or X iff Y. That is, if this relationship can be said to hold, for any individual who is in pain their C-fibers must be firing, and when C-fibers are firing in an individual, they are in pain. As is often the case, the “real” state of affairs turned out to be much more complicated (see Hardcastle, 1997, for an insightful discussion).

Token identity states that, at any given point in time, for a person being in a psychological state X, this state X is identical to some neurological state Y that instantiates it. However, it puts no restrictions on the consistency on this relationship: In the same person, the same psychological state X may be identical to a different neurological state Y2 at a different point in time. This position then, as Barrett notes, carries no real ontological weight and is better seen as an antidualist grounding of psychological states in physical reality. We then have two versions—one that puts a very strong restriction on the patterns of covariation across people, and the other that puts no restrictions on the nature of the identity relationship, only that there must be some identity theoretical relationship at any given time. We may then consider an alternative in between these two options, which we provisionally call type-token identity theory.

Type-token identity, in our interpretation, holds that *for every individual*, there exists a type identity theoretical relationship between a psychological state X and some neurological realization Y. That is, whenever I am in psychological state X, I am in neurological state Y (and vice versa), but this state Y may be quite different across people. Type-token identity then essentially formulates a type identical (reflective) model as we proposed in our original approach, but at the level of every individual (i.e., it may differ across individuals). For instance, take “depression,” a psychological construct that varies both across people and within people over time. If a reductionist description, for me, is not the same as for someone else, this means the group-level type identity (type-type) description, model will not fit. Measuring fluctuations of a suitable constellation of (P and N) indicators over time may then provide support for a unidimensional, reflective model of depression for each individual. Being depressed for me may mean, say, increased activity in brain regions A and B, and a decrease in hormone level C, but for a different person, the exact same state of depression (i.e., being equally depressed) may correspond to a different constellation of As, Bs, and Cs.

There are substantive and philosophical reasons to consider this interpretation of type-token identity as a viable candidate. First, as both Vul and Barrett note in their comment, and Bechtel and Mundale (1999) in previous work, if we go “down” the reductive ladder, any (type identical) model becomes token identical by necessity, simply because people do not have the same brains. We agree with this suggestion but think that the important question is whether *there is a level at all* for which, for instance, a reflective model can be said to hold. Here we part with Bechtel and Mundale (1999), who stated that if one chooses the correct granularity, “the mapping between them will be correspondingly systematic” (p. 202). We think that whether such a level (that achieves one-to-one mapping) exists for a given psychological construct is an open, empirical question that will in all likelihood be answered differently for different constructs. If there is an explanatory level at which type-type identity can be said to hold, then we would consider the type identity description more parsimonious and therefore preferable. If a type-type identity relationship does not exist for a particular psychological construct, that is, if people who can be

considered equivalent on some psychological dimension are consistently different along some neurological dimension, but there are observable regularities, type-token identity may be a good candidate.

Consider two people, as in Figure 5, who we could measure psychologically and neurologically with respect to some (latent) variable of interest, say, “depression.” We assume we can measure these people repeatedly and that depression is something that can vary within (over time) and between people (at any given time). It may turn out that a particular neurological property covaries with self-reported and observed levels in both individuals but that these neurological properties differ between them (cf. Barrett, 2006, p. 34 for a similar approach). Figure 5 shows how we can specify an identity theoretical model of a psychological construct for both individuals, with different parameter values. Intersubject variability can then explicitly be modeled, and individual differences in neurological measurement models may be described.

This would be in line with Frege’s (1956) remark, “He does not have my pain and I do not have his sympathy” (p. 300): Although it may be possible to construct

Figure 5. A type-token identity model. *Note.* Here we estimate the latent trait for two individuals by a time-series analysis (time-series not shown). In this way, we can estimate a model that incorporates neurological and psychological indicators for each individual. If we measure the same indicators for all individuals in the population, we can compare a random effects model (where you estimate parameters for every individual) to a model where the factor loadings are set to be equivalent for the whole population. A likelihood test can be used to assess whether the model improves enough to justify estimating the parameters for each individual.

an identity theoretical model concerning, say, pain or depression and its neurological realizations for every individual, it need not be identical for every individual to still provide valuable insights. This same structure may also hold for Berkman and Lieberman's treatment of the multiple realizability of psychological properties with respect to behavior: For each individual, an identity theoretical relationship between psychological properties and behavior may hold, but this may differ across people.

An empirical example of a study where type-token identity may be an appropriate perspective is research into selectively firing neurons or neural units. Consider, for instance, the case of "Jennifer Aniston cells" (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). "Jennifer Aniston cells" are a designation of a (hypothetical) neuron or constellations of neurons that fire only for extremely specialized stimuli, such as representations of Jennifer Aniston (or Halle Berry), invariant over the visual nature of these representation (i.e., these clusters of cells fired for both the linguistic stimulus "Jennifer Aniston" and a picture of her, but not picture of a similar-looking woman). These clusters of cells cannot be anatomically identified in such a manner that the description generalizes across individuals: They differ in location and structure across individuals. However, it may still be the case that all people, at least those with sufficient cultural exposure to the prior examples, display a constellation of cells that "behave" in this manner, that is, cells for which firing rate can be considered a measurement of the latent variable "perception of Jennifer Aniston." It may be possible to fit an identity theoretical model for each individual, where the latent variable (a dichotomous variable that represents "whether or not a person perceived a stimulus representing Jennifer Aniston") may be measured *either* by simply asking them *or* by measuring activity in this region/cluster of cells. If this description is empirically true, asking someone, "Did you see Jennifer Aniston?" may be an equally good, or valid, measurement of this latent variable as observing the spiking rate of this hypothetical cluster of cells. Similarly, in the previously discussed publication by Kanwisher et al. (1997), the data analysis yielded different (but similar) neuroscientific findings for the 15 participants in the experiment, possibly compatible with a type-token identity perspective.

We may consider identity theory as a guide to represent a hierarchical set of models. For certain psychological properties, it may be possible to single out a reflective identity theoretical model for a whole population, but for other properties (presumably, more fine-grained properties), a better approach may be a model at the individual level, or a supervenience approach. If there are neurological properties for which a group-level identity model holds, this is a stronger, more restrictive, and more parsimo-

nious (with fewer parameters) model, and as such is preferable.

Emergence and Mutualism

A final theory we consider for its empirical and conceptual implications is emergence. It is briefly mentioned by Bagozzi (this issue), and in more detail by Barrett (this issue). A general description of emergence is as follows: "Emergent phenomena are conceptualized as occurring on the macro level, in contrast to the micro-level components and processes out of which they arise" (Goldstein, 1999, p. 49). The brain can be seen as a prime example of a highly complex structure made up of basic building blocks that work together in complex ways. Strong emergence (Chalmers, 2006) is commonly viewed as the thesis that emergent phenomena, such as the capacity of the brain to engage in cognitive functions, have causal powers and an ontological status that cannot be fully reduced to its constituent components. For instance, consciousness can, it seems, not be explained on the basis of individual neurons (or the molecules that make up these neurons), only as an emergent feature of a complex system.

An emergent perspective on psychological properties may be appropriate, and in fact mathematically tractable, once we take into account the dimension of *time*. Consider intelligence, or *g*. A model of mutually interacting cognitive faculties has been proposed (van der Maas, Dolan, Grasman, Wicherts, Huizenga, & Raijmakers, 2006). This model shows how positive interactions of independent cognitive components over time may yield a higher order phenomenon (*g*, or the positive manifold) that is not reducible to individual cognitive components but requires knowledge of their interacting properties (in the description of Bechtel, 2008, their organizational properties). Expanding this model to include properties of the brain, we may envisage a model to include both measurements of these (independent) cognitive subsystems, and properties of the brain such as gray matter density or white matter connectivity for certain regions. Then, crucially, we would measure people (preferably children during development) through developmental time. In this way, we may be able to model and track how higher order phenomena arise from the interaction of lower order components, and so distinguish between hypotheses that, with cross-sectional data, cannot be differentiated (cf. Barrett, this issue). That such time series modeling of brain measurements through time may yield counterintuitive results was illustrated in research done by Shaw et al. (2006). They showed that the nature of the relationship between certain neurological indicators (in this case, cortical thickness) and psychological measurement of general

intelligence changed over time: the correlation between cortical thickness and IQ was negative for the youngest children but was positive for the same group at a later developmental stage. It is these types of developmental phenomena that fit in well with an emergent perspective on higher order properties. We may then be able to specify a measurement model for such developing systems that represent the evolving interactions between psychological and neurological indicators. In this manner, it may be possible to refine both our measurement models and mechanistic explanations of these phenomena.

However, we should state this is no trivial task: A good measurement model does not necessarily include indicators that are most insightful about the underlying mechanisms, and vice versa. For instance, if we have two component parts that are necessary for a mechanism to “work,” (activity of) these components may not covary with observable variation in function. This is precisely why results from fMRI should be interpreted with caution with respect to underlying mechanisms: Neural activity that may be essential for a particular psychological function (e.g., working memory) but does not covary with various experimental conditions in the task (e.g., high or low working memory demands) will generally not be detected by means of traditional General Linear Model (GLM) designs, despite the fact that some of this undetected neural activity may be of importance for the psychological function of interest.

Future Developments

So how should we move forward? First, we think the models proposed by the commenters and in our work could, and should, be applied to a selection of current empirical paradigms. It will be of great interest to see which types of models best represent the relationship between psychological phenomena and properties of the brain, and whether this is different for different types of psychological phenomena. One can imagine a “scale” of models ranging from strict to lenient, where we speculate that certain lower psychological phenomena (e.g., depth perception) may fit more statistically stringent models than higher order phenomena (e.g., intelligence). This scale of model fit, ranging from a tight mapping between neural indicators and psychological indicators to a much looser mapping, may be considered akin to Bickle’s (1998) notion of “smooth” and “bumpy” reduction. As an example, fitting a type-type identity theoretical model to a set of psychological and neurological measurements could be considered an example of relatively smooth reduction.

Second, we should continue to look out for, and where necessary attempt to develop, new psychometric models. Although the reflective and formative mod-

els we discuss have a relatively long history, there are plenty of developments in the field of psychometric modeling. For instance, a whole special issue of the *Journal of Business Research* was devoted just to the interpretation and further development of formative models, such as the proposals by Diamantopoulos, Riefler, and Roth (2008) to combine various aspects of reflective and formative models within the same measurement model. Other noteworthy developments are latent variables models specifically aimed at addressing time-series data sets such as we discussed in the type-token model example in Figure 5 (Hamaker, Nesselrode, & Molenaar, 2007), and developments that have made Bayesian analysis of latent variable models possible (e.g., Congdon, 2003, chap. 8).

Of course, the field of cognitive neuroscience itself is not sitting still on the topic of structuring our current knowledge and furthering new approaches. For instance, several commenters note the Cognitive Ontology project of Poldrack (e.g., Poldrack, 2006). This project aims to structure the knowledge concerning psychological constructs and the known “neural correlates” so that we may gain progressive insight concerning the current neuroscientific findings with respect to psychological constructs. This is a fascinating project that could be extended with measurement models. For instance, for a given construct present in the cognitive ontology database, what is our current best measurement model to model this constructs? If someone were to look up “cognitive control” and integrate it into a new fMRI design, what is currently our best measurement theoretical description of indicators? These are but some ways in which the psychometric approach can be integrated into current research lines.

Conclusion

The realization that for the field of cognitive neuroscience to grow from an “adolescent field” (Kriegeskorte, 2010) into an adult field it needs to formalize its models has been steadily gaining track (e.g., Just, Cherkassky, Aryal, & Mitchell, 2010; Mars, Shea, Kolling, & Rushworth, 2010; Miller, 2010; Yarkoni, Poldrack, Van Essen, & Wager, 2010). Cognitive neuroscience is an incredibly complex topic, but significant advances have been made, and will be made.

The commenters have done much work in furthering, both conceptually and empirically, the suggestions we made in our target article. Although the reduction problem is indeed “not solely a measurement problem” (Bagozzi, this issue, p. 98), we hope we have shown the benefits of treating it as such. We have argued in this reply that the extent to which the brain is informative with respect to specific psychological properties should be considered an open question. Significant advances have been made, and increasingly psychological

constructs and common neurological measurements are being refined based upon empirical evidence. For the time being, we will take Barrett's (this issue) advice: "For the present, however, it makes sense to forge ahead . . ." (p. 120). The insightful nature of the comments to our target article suggest that reframing the relationship between neurological and psychological measurements by means of formal models is a fruitful strategy that may help to give us a clearer view of the nebulous relationship between the mind and the brain.

Acknowledgments

This article is original. None of the materials have been published elsewhere. We thank Angelique Cramer, Petry Kievit-Tyson, and Anne-Laura van Harmelen for valuable comments on earlier versions of this article.

Note

Address correspondence to Rogier A. Kievit, University of Amsterdam, Department of Psychological Methods, Roetersstraat 15, 1018WB Amsterdam, the Netherlands. E-mail: r.a.kievit@uva.nl

References

- Aydede, M. (2000). On the type/token relation of mental representations. *Facta Philosophica*, 2, 23–49.
- Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, 1, 28–58.
- Barrett, L. F. (2009). The future of psychology: Connecting mind to brain. *Perspectives on Psychological Science*, 4, 326–339.
- Bassett, D., & Bullmore, E. (2006). Small-worlds brain networks. *Neuroscientist*, 12, 512–523.
- Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York, NY: Routledge/Taylor & Francis.
- Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science*, 66, 175–207.
- Bickle, J. (1998). *Psychoneural reduction. The new wave*. Cambridge, MA: MIT Press.
- Bock, R. D., & Kolakowski, D. F. (1973). Further evidence of sex-linked major gene influence on human spatial visualizing ability. *American Journal of Human Genetics*, 25, 1–14.
- Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, 97, 303–352.
- Chalmers, D. J. (2006). *Strong and weak emergence*. In P. Clayton & P. Davies (Eds.), *The re-emergence of emergence* (pp. 244–256). Oxford, UK: Oxford University Press.
- Chiao, J. Y., & Cheon, B. K. (2010). The weirdest brains in the world. *Behavioral and Brain Sciences*, 33, 88–90.
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78, 67–90.
- Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge, UK: Cambridge University Press.
- Coan, J. A. (2010). Emergent ghosts of the emotion machine. *Emotion Review*, 2, 274–285.
- Congdon, P. (2003). *Applied Bayesian modelling*. Chichester, UK: Wiley.
- Dennett, D. C. (1991). Real patterns. *Journal of Philosophy*, 88, 27–51.
- Diamantopoulos, A. (2006). The error term in formative measurement models: interpretation and modeling Implications. *Journal of Modelling in Management*, 1, 7–17.
- Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008) Advancing formative measurement models. *Journal of Business Research*, 61, 1203–121.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174.
- Fischer, H., Wright, C. I., Whalen, P. J., McInerney, S. C., Shin, L. M., & Rauch, S. R. (2003). Brain habituation during repeated exposure to fearful and neutral faces: A functional MRI study. *Brain Research Bulletin*, 59, 387–392.
- Fodor, J. A. (1974). Special sciences (Or: The disunity of science as a working hypothesis). *Synthese*, 28, 97–115.
- Frege, G. (1956). The thought: A logical inquiry. *Mind, New Series*, 65, 289–311.
- Goldstein, J. (1999). Emergence as a construct: History and issues. *Emergence*, 1, 49–72.
- Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, 40, 303–329.
- Hamaker, E. L., Nesselrode, J. R., & Molenaar, P. C. M. (2007). The integrated trait- state model. *Journal of Research in Personality*, 41, 295–315.
- Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21, 1639–1651.
- Hardcastle, V. G. (1997). When a pain is not. *Journal of Philosophy*, 94, 381–390.
- Heils, A., Teufel, A., Petri, S., Stober, G., Riederer, P., Bengel, D., et al. (1996). Allelic variation of human serotonin transporter gene expression. *Journal of Neurochemistry*, 66, 2621–2624.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960.
- Horgan, T. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind*, 102, 555–586.
- Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS One*, 5, e8622.
- 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.
- Karlin S. (1983, April 20). *11th R.A. Fisher Memorial Lecture, Royal Society*.
- Kendler, K.S. (2005). "A gene for . . .": The nature of gene action in psychiatric disorders. *American Journal of Psychiatry*, 162, 1243–1252.
- Kievit, R. A., van Rooijen, H., Wicherts, J., Scholte, H. S., Waldorp, L. J., & Borsboom, D. (2011). *If neuro g is the answer, what is the question?* Manuscript submitted for publication.
- Kim, J. (1987). "Strong" and "global" supervenience revisited. *Philosophy and Phenomenological Research*, 48, 315–326.
- Kim, J. (2005). *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press.
- Kriegeskorte, N. (2010). Interpreting brain images: Reflections on an adolescent field. *Trends in Cognitive Sciences*, 14, 475–476.
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, 12, 535–540.

- Krukow, K., Nielsen, M. & Sassone, V. (2008). Trust models in ubiquitous computing. *Philosophical Transactions of the Royal Society*, 366, 3781–3793.
- Lamme, V. A. F. (2004). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10, 494–502.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23, 571–579.
- Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70, 556–567.
- Lewis, D. (1994). Human supervenience debugged. *Mind, New Series*, 103, 473–490.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Augelman, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412, 150–157.
- MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, 55, 95–107.
- McCauley, R. N., & Bechtel, W. (2001). Explanatory pluralism and heuristic identity theory. *Theory and Psychology*, 11, 736–760.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly*, 2, 245–264.
- Mars, R. B., Shea, N. J., Kolling, N., & Rushworth, M. F. S. (2010). Model-based analyses: Promises, pitfalls, and example applications to the study of cognitive control. *The Quarterly Journal of Experimental Psychology*. Advance online publication.
- McCrae, R. R., & Costa, P. T., Jr. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36, 587–596.
- McMullin, E. (1983). Values in science. In P. Asquith & T. Nickles (Eds.), *Philosophy of Science Association 1982, Vol. 2* (pp. 3–25). East Lansing, MI: Philosophy of Science Association.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543.
- Miller, G. A. (2010). Mistreating psychology in the decades of the brain. *Perspectives on Psychological Science*, 5, 716–743.
- Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality*, 21, 549–587.
- Pigliucci, M. (2007). Do we need an extended evolutionary synthesis? *Evolution*, 61, 2743–2749.
- Place, U. T. (1999). Token- versus type-identity physicalism. *Anthropology and Philosophy*, 3, 21–31.
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10, 59–63.
- Poline, J.-B., Thirion, B., Roche A., & Meriaux, S. (2010). *Inter-subject variability in fMRI data: Causes, consequences and related analysis strategies*. In S. J. Hanson & M. Bunzl (Eds.), *Foundational issues in human brain mapping* (pp. 173–191). Cambridge, MA: MIT Press.
- Posthuma, D., Baaré, W. F. C., Hulshoff Pol, H. E., Kahn, R. S., Boomsma, D. I., & De Geus, E. J. C. (2003). Genetic correlations between brain volumes and the WAIS–III dimensions of verbal comprehension, working memory, perceptual organization, and processing speed. *Twin Research*, 6, 131–139.
- Psillos, S. (2004). A glimpse of the secret connexion: Harmonizing mechanisms with counterfactuals. *Perspectives on Science*, 12, 288–319.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435, 1102–1107.
- Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.
- Rowlands, M. (1992). Externalism and token-token identity. *Philosophia*, 22, 359–375.
- Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81, 961–962.
- Schleim, S., & Roiser, J. P. (2009). fMRI in translation: The challenges facing real-world applications. *Frontiers in Human Neuroscience*, 3, 63.
- Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., & Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, 440, 676–679.
- van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience*, 29, 7619–7624.
- van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (in press). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*. doi: 10.1037/a0022749.
- van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P., Wicherts, J. M., Huizenga, H. M. & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113, 842–861.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440–442.
- Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science*, 69, 366–377.
- Yarkoni, T., Poldrack, R. A., Van Essen, D. C., & Wager, T. D. (2010). Cognitive neuroscience 2.0: Building a cumulative science of human brain function. *Trends in Cognitive Sciences*, 14, 489–496.