Cognitive neuroscience involves the simultaneous analysis of behavioral and neurological data. Common practice in cognitive neuroscience, however, is to limit analyses to the inspection of descriptive measures of association (e.g., correlation coefficients). This practice, often combined with little more than an implicit theoretical stance, fails to address the relationship between neurological and behavioral measures explicitly. This article argues that the reduction problem, in essence, is a measurement problem. As such, it should be solved by using psychometric techniques and models. We show that two influential philosophical theories on this relationship, identity theory and supervenience theory, can be easily translated into psychometric models. Upon such translation, they make explicit hypotheses based on sound theoretical and statistical foundations, which renders them empirically testable. We examine these models, show how they can elucidate our conceptual framework, and examine how they may be used to study foundational questions in cognitive neuroscience. We illustrate these principles by applying them to the relation between personality test scores, intelligence tests, and neurological measures.

There is nothing more practical than a good theory.
— Lewin (1951)

One of the hallmark neuroscientific findings of the 20th century is the discovery of the retinotopic representation of early visual areas (e.g., Hubel & Wiesel, 1968; Tootell, Switkes, Silverman & Hamilton, 1988). That is, activation patterns in the occipital lobe show striking structural similarity to visually presented geometric patterns. Such findings, originally only possible in animal research, have been replicated in humans in more indirect form. For instance, Miyawaki et al. (2008) showed how basic visual stimuli (including letters) can be decoded from brain activity with high accuracy (>90%), based upon weighted linear combinations of voxel activation patterns. For such low-level perceptual processes, it seems plausible to consider the observation of activity patterns in early visual areas as a measurement of what particular stimulus is presented to a particular subject. However, the measurement theoretical relationship is not always so clear. Consider the following example: You are invited to a job interview for a high-status position. Shortly after being seated, the interviewer takes out a tape measure and starts measuring your skull. Upon enquiring what is going on, the interviewer tells you he just “measured your intelligence.” In response to your protesting that such a procedure does nothing of the sort, the interviewer shows you a list of high-profile journal articles that report a moderate but consistent correlation between brain volume and IQ (e.g., McDaniel, 2005; Posthuma et al., 2002). You may believe that such a procedure does not measure intelligence, but this appears to run counter to the view in
cognitive neuroscience\textsuperscript{1} that physiological measures may serve as measures of psychological attributes. We later return to the empirical formalization of this question.

What is the essential difference between these two situations? Both take information about the brain to predict a certain (psychological) property, and both are based on statistically significant measures of association, but at the same time they seem quite distinct. It seems thoroughly unclear how to resolve this issue. This raises two questions: of how cognitive neuroscientists actually represent the relationship between the two classes of measures, and what presentation would justify the interpretation of neurological measures as representing psychological attributes.

The general practice in cognitive neuroscience is to limit statistical analyses to the study of descriptive measures of association (e.g., correlation coefficients). In fact, some authors have argued that cognitive neuroscience is by its very nature correlational (e.g., Jung & Haier, 2007, p. 148). However, this would leave open important questions: What is the precise relationship between these two classes of measurement? Does one measured property cause the other? Or is it the other way around? Do the different kinds of data really represent measurements of the same thing? Many publications implicitly embrace one of these options, possibly because there simply is no “value-free” way in which to describe the relationship between behavioral measurements and neurological measurements—unless, perhaps, if one is satisfied with the conclusion that “they both just happened.” Certainly it is desirable (if not tempting) to attach some theoretical interpretation to the established empirical relationship between psychological-behavioral and neurological measures. However, the mere inspection of correlation coefficients provides no sound basis for deciding between different theoretical interpretations.


Conceptual problems in reductive psychological science have not gone unnoticed. Several researchers have taken on theoretical, statistical, and scientific issues concerning reductionism and reductive psychological science. For instance, Bennett and Hacker suggested that the vocabulary employed in neuroscientific studies is conceptually flawed. One of the issues they raised is the “mereological fallacy,” or “assigning to a part what can only be assigned to a whole” (Bennett & Hacker, 2003, p. 68). They identified this fallacy in statements such as “the frontal lobe engages in executive functioning.” They argued that this practice is philosophically misguided and reflects a conceptual problem within reductive neurological science. Ross and Spurrett (2004) argued that functionalist cognitive psychology requires a solid metaphysical underpinning of the conceptual and scientific foundations if it is to function as an autonomous field of scientific inquiry. Other researchers (Fodor, 1974; Gold & Stoljar, 1999; Nagel, 1961) have examined the philosophical foundations of reductionism and explicated the requirements necessary for reductionist claims. Recent efforts have examined whether the ontology of psychological categories is suitable for reductive analysis and argued that an approach in terms of psychological primitives may be more appropriate (Barrett, 2009). Criticism of reductive studies has not been purely philosophical. In a controversial article, Vul, Harris, Winkielman, and Pashler (2009) argued that a large number of claims in social neuroscience studies are overstated and that overly liberal methodology has resulted in unrealistically high correlations between physiological and behavioral measurements (but see also associated comments to Vul et al., 2009).

These publications have focused largely on what conclusions are not permissible, methodologies that should not be used, and philosophical claims that can not be made. The aim of the current article is to address the criticisms raised by the aforementioned authors by providing conceptual and statistical tools that may elucidate the type of claims that we can make in reductive science and developing the requirements such claims should satisfy.

\textsuperscript{1}We refer to the discipline here as cognitive neuroscience, as it is the broadest and most common name for the concurrent study of psychological behavior and physiological properties. However, we do not aim to restrict our perspective to merely cognitive phenomena such as attention, memory, or intelligence: The issues we raise are equally of interest for fields such as social neuroscience or affective neuroscience. Wherever we state cognitive neuroscience, we mean to encompass such more specific branches.
Cognitive neuroscience typically attempts to establish the relationships between at least two distinct explanatory levels, namely, the neurological and psychological level (Oppenheimer & Putnam, 1958). As such, it has drawn much attention from philosophers, who have articulated and analyzed many theoretical positions regarding the relations between the two levels of analysis (e.g., Churchland, 1981, 1985; Kim, 1984; Lewis, 1966; Putnam, 1973). Several philosophers recently developed perspectives on reduction, seeking to integrate certain developments in, for example, molecular neuroscience (e.g., the “New Wave Reductionism” promoted by Bickle, 1998). It would seem that if such positions could be translated into statistical models that are testable given the data that cognitive neuroscientists commonly have at their disposal, the theories articulated in the philosophy of mind could serve as a means to conceptually organize and guide the analysis of neurological and behavioral data. That is, if it were possible to find a statistical model representation of, say, the basic assumption that the property measured by means of fMRI recordings actually is the same as the property measured through a set of cognitive tasks or questionnaire items (i.e., identity theory; Lewis, 1966), then both the philosopher of mind and the empirical researcher in cognitive neuroscience would benefit: the philosopher of mind, because there would exist a means to empirically test theories that have hitherto been regarded as being speculative metaphysics at best, and the empirical researcher, as this could provide statistical tests of interpretations of the data that go well beyond the speculative interpretations of correlations that currently pervade the literature.

How could statistical models be of help to the empirical researcher in cognitive neuroscience? Recall that, in this area of research, one typically aims to build connections between measures related to behavior, psychological attributes, and processes, on the one hand, and the (relative) activity and physiological characteristics of the brain, on the other hand. In psychometrics, we can represent such diverging classes of measurement in a single measurement model. The central idea of this article is that by varying the way in which a theoretical attribute relates to the observations, models can be built that allow for a more detailed investigation of the relation between neurological and psychological measurements than are in use to date. This article proposes that modeling techniques suited to this purpose need not be developed for this purpose, because they already exist. These mathematically tractable models with known statistical properties, developed largely in the discipline of psychometrics, can map theoretical positions about the relationship between brain and behavioral measurements as developed in the philosophy of mind in impressive detail. We argue that the statistical formalization of theoretical positions is both possible and desirable, and we offer the empirical and conceptual tools to do so. Perhaps most important, such formalizations make clear that the reduction problem is, in essence, not just a substantive or philosophical problem but a standard measurement problem that can be attacked by using standard measurement models of psychometrics. However, such models have been scarcely applied in cognitive neuroscience. From this perspective, therefore, it seems as if most empirical work has, instead of solving the measurement problem, largely circumvented it.

The structure of this article is as follows. We first define the two classes of measurement under study. Subsequently, we examine two important theories from the philosophy of mind literature that explicitly treat the relationship between these higher and lower order properties, namely, identity theory and supervenience theory. In addition, we introduce two psychometric models that may be used to represent these theoretical positions. Finally, we illustrate these ideas by applying both models to datasets examining the relationship between a personality dimension and intelligence on the one hand to physiological properties of the brain on the other hand.

Two Types of Data

In the models we discuss next, we distinguish between the two classes of data that feature in most cognitive neuroscientific studies. First, we refer to data that pertain to psychological attributes or mental processes as P-indicators. These include psychological measurements, such as “solving puzzle x,” “choosing answer c,” or “the number of objects retained in working memory.” Second, we refer to data that pertain to neurological processes or characteristics as N-indicators. These may include data such as electrical measures of cortical activity (EEG), speed of processing measurements, blood oxygenation level-dependent (BOLD) signals, as well as physiological indicators such as gray matter density, brain volume, or neurotransmitter levels. The psychological indicators are indexed to denote either different questions on a test (P1 is one question, P2 another) or different types of measurement (P1 is an IQ score, P2 a reaction time test). Neurological indicators are indexed to denote, for example, different regions of the brain (e.g., N1 is a BOLD measurement of the posterior parietal region, N2 of the amygdala), or different types of physiological variables (N1 is gray matter density, N2 is neural processing speed).

For example, in a cognitive neuroscientific study of empathy, psychological measurements of empathy could include P-indicators such as questionnaires, self-reports, or behavioral assessments. In contrast, neurological measurements would include N-indicators such as the level of BOLD-activation in certain cortical regions in response to seeing another person suffer (see Decety & Jackson, 2004, for a review of the
Supervenience provides an alternative way of conceptualizing the relation between psychological and neurological measurements. Different interpretations of supervenience have been formulated in relation to a wide range of philosophical topics (Collier, 1988; Hare, 1952; Horgan, 1993). Historically, the concept arose from attempts to ground the properties of higher level concepts such as beauty, morality, and consciousness in their lower order realizations. The definition of supervenience is as follows: A property X can be said to supervene on lower order properties Y if there cannot be X-differences without Y-differences. Thus, the presence of Y-differences is a necessary (but insufficient) condition for the presence of X-differences. This relation of necessity is a sufficient condition for calling the relation one of supervenience. Consider, for example, the attribute of being morally good. Under supervenience theory, two people cannot differ in terms of morality (X) without being different on lower order Y attributes (e.g., behavioral ones; not stealing, cheating, donating money to charity, etc). Equivalently, if there are no differences in the lower order attributes (Y, or behavioral attributes), then there are necessarily no differences in the higher order attribute (X, or morality). This is the sense in which morality supervenes on its lower order attributes. Properties such as morality and beauty are “along for the ride,” so to speak: They supervene on lower order properties that do not necessarily share all the characteristics that relate to the supervenient property. The atoms that make up the Mona Lisa are not beautiful, and neurons are not neurotic: Such higher order properties supervene on the lower order properties in a causally asymmetric manner.

The philosophical details of supervenience are still the subject of theoretical perspectives and debates. Its most vocal advocate in the realm of psychology has been Jaegwon Kim. His supervenience perspective on psychology (Kim, 1982, 1984, 1985) defines psychological attributes as supervenient on neurological realizations. That is, psychological attributes are completely determined by, or realized in, their neurological constituents. Supervenience has been the topic of various recent debates on specific alternative interpretations of the concept, varying in terms of modal strength and necessity (Horgan, 1993; R. J. Howell, 2009). Although these are of interest in and of themselves, a comprehensive discussion would lead us too far astray from our current aim. For sake of parsimony, we adopt Kim’s more traditional definition of strong supervenience. Kim defined the supervenient status of higher and lower level properties A and B, respectively, as follows: “Necessarily, for any x and y, if x and y share all properties in B, then x and y share all properties in A—that is, indiscernibility in B entails indiscernibility in A” (Kim, 1987, p. 315). The relationship of supervenience is asymmetric, as neurological states or structures can differ, whereas the higher order property remains the same (because lower order differences are necessary, but not sufficient, for higher order differences).

This implies that supervenience allows for multiple realizability (Putnam, 1980); several different combinations of N-realizations may lead to the same (value of the) psychological attribute. Because of this asymmetry, authors such as Kim give causal priority to the lower order realizations: The neurological indicators are considered to determine the causal properties of the system completely. Supervenience is consistent with a many-to-one mapping of the lower to the higher order properties, but not with an isomorphism (which would hold if P-realizations determine N-realizations).

Identity Theory

The thesis of identity theory was proposed in several forms throughout the latter half of the 20th century. It has its roots in seminal publications such as those of Place (1956) and Smart (1959). In its most commonly accepted interpretation, as described in Lewis (1966), identity theory holds that psychological processes and attributes are identical to their neurological realizations.

The attractiveness of identity theory lies in the relatively nonproblematic assignment of causal powers to mental events. Because a mental event or state is identical to a (particular) neural realization at any given time, it has the same causal powers as the neurological state that realizes it. This implies that in a cognitive neuroscientific study of a particular psychological attribute, one is essentially measuring the same attribute using two different measurements. The P- and N-indicators therefore have a common referent. This conceptualization paints a thoroughly realist picture of psychological attributes, in which the reality of these attributes is grounded in their physical realization.

Supervenience

Supervenience provides an alternative way of conceptualizing the relation between psychological and neurological measurements. Different interpretations of supervenience have been formulated in relation to a wide range of philosophical topics (Collier, 1988; Hare, 1952; Horgan, 1993). Historically, the concept arose from attempts to ground the properties of higher level concepts such as beauty, morality, and consciousness in their lower order realizations. The definition of supervenience is as follows: A property X can be said to supervene on lower order properties Y if there cannot be X-differences without Y-differences. Thus, the presence of Y-differences is a necessary (but insufficient) condition for the presence of X-differences. This relation of necessity is a sufficient condition for calling the

---

2 A similar position can be found in D. Davidson (1980, p. 111).
hold if all relations between instances of the lower order terms are preserved in the higher order relations), and therefore precludes identity.

To illustrate this, consider the following transaction. If John gives Jane $5 (higher order process), then that means that John has either given Jane a $5 bill, has handed her the equivalent sum in coins, or has electronically transferred $5 to Jane’s bank account (lower order processes). Thus, the entire class of these lower order processes maps onto the same higher order process. If we know that John gave Jane $5, we can therefore infer that he performed one of the actions in the corresponding lower order class. However, we cannot determine which of these actions he performed (no isomorphism). It is evident that an identity theory perspective on such a monetary transaction is questionable: John giving Jane $5 cannot simultaneously be identical to writing a cheque and to handing over a $5 bill. We now show how such restrictions and theoretical considerations can be translated to and mapped on psychometric models. To do so, we must first examine the basic properties of the models that we consider.

**Psychometrics**

Psychometrics is concerned with the theoretical and technical development of measurement procedures and statistical inference techniques. One of the techniques, developed in tandem with psychometric theory, is structural equation modeling (SEM). SEM consists of both a graphical and a (equivalent) linear mathematical representation of the hypothesized causal directions and statistical associations between measured and latent variables. Such representations imply a specific covariance structure, which may be tested given appropriate data. Specifically, one can evaluate whether the observed covariance matrix is consistent with the covariance structure associated with the specified linear relationships. For a thorough introduction to SEM with latent variables, see Bollen (1989).

In SEM, there are two broad classes of model specification that we consider in detail, namely, **formative** and **reflective** models (Bagozzi, 2007; Bollen & Lennox, 1991; Edwards & Bagozzi, 2000). Both classes model relationships between observed variables and latent variables. Here “observed variables” refer to the variables as they appear in a data file, and “latent variables” refer to variables that are not directly observable, so that their values can only be estimated indirectly (Bollen, 2002; Borsboom, 2008). Many of the properties central in psychological science (e.g., intelligence, personality, working memory capacity) cannot be determined with certainty from the data and are therefore properly conceived of as latent variables.

Formative and reflective models provide two ways of connecting a theoretical attribute, as targeted by a measurement procedure, to the observations. We discuss the conceptual difference between these two models in relation to the distinction between identity theory and supervenience theory. We present the models using standard SEM notation (Jöreskog & Sörbom, 1996). As mentioned, SEM permits the specification of linear relations between the observed and latent variables as implied by theoretical considerations, and the evaluation of the degree to which the observed covariance structure is consistent with that implied by the theoretical relations. The models can allow for either tentative confirmation, in the sense that they fit the data, or rejection, in the sense that they can be overspecified or display poor fit. Thus, these models are amenable to empirical tests.

**Reflective Models**

The most common measurement model in psychology is called the reflective model. Instances of the model include Item Response Theory models (Embreton & Reise, 2000), such as the models of Rasch (1960) and Birnbaum (1968), and, most relevant to the present article, the linear factor model (Jöreskog, 1971; Lawley & Maxwell, 1963; Mellenbergh, 1994). In reflective models, latent variables are seen as the underlying cause of variability on the measurable indicators (Bollen, 2002; Bollen & Lennox, 1991; Borsboom, Mellenbergh, & van Heerden, 2003; Edwards & Bagozzi, 2000). In other words, the hypothesized causal direction runs from the latent attribute to the measurable indicators. The various measurable indicators are seen as reflecting the underlying attributes. Perhaps the most common example of a reflectively measured attribute is intelligence. The conceptualization of intelligence posits a factor $g$ that refers to the common cause of variability on intelligence test questions or subtests (Glymour, 1998; Jensen, 1998).

A reflective model of $g$ is given in Figure 1. In the figure, three indicators (e.g., IQ test items or subtest scores) are conceptualized as measurements of a single underlying attribute (this is a simple, nonhierarchical model of $g$, chosen for illustrative purposes). Indicators of a reflectively measured latent variable should (after appropriate recoding) intercorrelate positively, capture the range of effects the latent variable can have, and be acceptably reliable (i.e., be characterized by acceptable levels of measurement error). In addition, in correctly specified reflective models, latent variables should be referentially stable. That is to say that the addition or deletion of an indicator may alter the accuracy by which the attribute is measured but not the nature of the attribute (latent variable) itself. With regard to the measurement of $g$, Spearman called this characteristic indifference of the indicators (Spearman, as cited in Jensen, 1998). Thus, the indicators are exchangeable in the sense that an exchange possibly
is based on the idea that the indicators determine the latent attribute, rather than the other way around. With respect to SES, this seems to be a plausible model. For instance, you do not get a raise because your SES level goes up; rather, your SES level goes up because you get a raise.

It is often argued that indicators in formative models should capture different aspects of the formative attribute and should not be too strongly related (Bollen, 1984; Diamantopoulos & Siguaw, 2006). The latent attribute in such a model is represented as the weighted sum of different indicators that together predict a relevant phenomenon. An important theoretical characteristic of this model is that the latent attribute is defined by the choice of predictors. Thus, in contrast to the reflective model, a change of predictors implies a change in the nature of the attribute. In addition, in many circumstances the theoretical attribute is referentially unstable because the weights of the connections between the observations and latent variable are usually constructed to maximize the prediction of external criteria. That is to say, the value of the latent variable for a given person may change from one study to the next, if the predicted criterion changes (Bollen, 2002, 2007; Burt, 1976; R. D. Howell et al., 2007).

Empirical Testability of Models

A crucial property of formative and reflective models is that they are testable, that is, they can be empirically corroborated or refuted, because the models impose restrictions on the joint probability distribution of the observations. Therefore, the support affects measurement properties such as precision but not the meaning of the attribute of interest. In a reflective model, observables are indicators of a common theoretical attribute, in the same way that a set of differently constructed thermometers are indicators of a common attribute, namely, temperature. Thus, it is assumed that the indicators measure the same thing. This implies that the latent variable or attributes exists independently of the model specification, at least with respect to the particular items used to measure it (Borsboom et al., 2003). Of course, positing a reflective model does not guarantee the existence of purported latent variables: Rather, the adoption of such a model generally carries with it a nontrivial ontological stance with regard to the latent variable.

Formative Models

Formative models express the relationship between theoretical attributes and observations in terms of a regression function in which the theoretical attribute features as the dependent variable and the observed variables as predictors. This is compatible with a conceptualization of the theoretical attribute (latent variable) as being in some way causally dependent on its indicators.

A common example of a formatively measured latent variable is socioeconomic status (SES), where the SES score for a given person is conceived of as a weighted sumscore of the measured variables, such as income and education level (R. D. Howell, Breivik, & Wilcox, 2007; Knesebeck, Lüschen, Cockerham, & Siegrist, 2003). Figure 2 depicts a path diagram of the formative model of SES. The three X indicators each contribute, with a certain weight, to the sumscore of the attribute SES. The Xs in this example could be income, education, or other variables deemed relevant to the estimation of SES. The structure of the model

![Figure 2. A formative model of socioeconomic status (SES). Note. The attribute, SES, is determined by the measured indicators. X = observed variable; ζ = residual term; η = weighted sumscore; the weights denoted are by γ.](image-url)
for a given specification of the underlying structure can be assessed by means of standard statistical tests and model-fitting methods. Many fit indices have been developed for the evaluation of the fit of SEM models (Hu & Bentler, 1999; Schermelleh-Engel, Moosbrugger, & Müller, 2003). Generally, fit indices are based on the discrepancy between the covariance structure implied by the specified model and the covariance structure, as observed in the data.

Commonly used fit indices are the chi-square for goodness-of-fit test, the root mean square error of approximation (RMSEA), and the comparative fit index (CFI). See Hu and Bentler (1999) and Schermelleh-Engel et al. (2003) for discussions of cutoff criteria for various fit indices for varying sample sizes and model complexity. A discussion of the details of model selection is beyond the scope of this article. The main point is that such models can be fitted to empirical data and that this yields well-developed quantifications of the adequacy of the model. For detailed considerations of model specification and fitting procedures, an extensive and active area of psychometric literature focuses on the optimal manner in which to examine model fit and model selection (R. D. Howell et al., 2007; Jarvis, Mackenzie, & Podsakoff, 2003; Myung & Pitt, 1997; Pitt, Myung, & Zhang, 2002, Waldorp, Grasman, & Huizenga, 2006), parameter estimation (Diamantopoulos & Siguaw, 2006; Myung, 2006), stability over time (Hamaker, Nesselroade, & Molenaar, 2007; Van Buuren, 1997), and issues such as interpretational confounding (Bollen, 2007; R. D. Howell et al., 2007). Given that we have many tools to determine the (relative) adequacy of our specified models, we now turn to the more relevant issue of how the theoretical positions discussed earlier may be mapped onto reflective and formative models.

Mapping of Psychometrics on Theory of Mind

We first examine identity theory, the theoretical position that at a given time psychological and neurological properties of measurements reflect the same attribute. This implies that both P- and N-indicators have a common underlying cause, namely, the true state of the latent variable. This is consistent with the reflective model, because that model views variability of the underlying attribute as the cause of variability in both P- and N-indicator values. Therefore, when measuring brain activity and psychological behaviors related to a particular phenomenon such as intelligence, one is essentially measuring the same thing. Figure 3 shows how variation in the latent attribute (e.g., a subject’s level of intelligence, or g) is the common cause of variation in both P-indicators (e.g., "giving the correct answer to a certain IQ-test question") and N-indicators (e.g., "increased activity in the dorsolateral prefrontal cortex"). If the reflective model of intelligence is correct, then the latent variable represents the actual value of g, which can be estimated in the same manner by both P- and N-indicators.

Therefore, P- and N-indicators can be said to be on equal empirical footing in that they are both assumed to be imperfect reflections of the true state of the underlying attribute. Identity theory is concordant with a realist perspective of psychological science, in the sense that it considers psychological attributes to be the underlying cause of variability of measurable indicators. The reflective model furnishes a psychometric implementation of identity theory: Both the conceptual and the psychometric model assume a singular underlying cause that can be measured by two methods. The

Given the exact formulation as a SEM, one should construe this to mean that variability in the underlying attribute causes variability in both the P- and the N-indicators.
expected values of measurements within this model can be expressed as a function of the value of the latent attribute and the parameter that expresses the strength of the relationship between attribute and indicator. As such, it can be tested in the same way as psychometric models are usually tested. Thus, the reflective model can be used to provide an empirical test of the identity hypothesis.

The conceptual advantage of the reflective model is that it allows for a substantive interpretation of both classes of measurement by equating the psychometric status of neurological and psychological indicators. For example, some scientists argue that psychological concepts or processes are best measured by psychological measurements, whereas others maintain that neurological measures are more precise or insightful (e.g., the process or concept of consciousness; cf. Lamme, 2006). This dissension concerning the merits of neurological and psychological measurements in measuring a psychological attribute seems coherent only from an identity theoretical perspective. A debate on the relative merits of two methods of measurement requires that the object of measurement be the same. This allows one to gauge the relative measurement precision of neurological and psychological indicators. At the same time, it allows for a comprehensible interpretation of both types of psychological research: A (non-neuroscientific) psychologist may acknowledge that corroborating evidence can be gained by the neurological approach (the same applies to the cognitive neuroscientist vis-à-vis psychometric data). Identity theory and reflective models view reductive psychological science as an integrated attempt to derive the best measure of the underlying attributes of interest. Such mutually insightful scientific interaction is in line with Heuristic Identity Theory (McCauley & Bechtel, 2001), which argues that simultaneous scientific study of two distinct explanatory levels from an identity theoretical perspective. A debate on the relative merits of two methods of measurement requires that the object of measurement be the same. This allows one to gauge the relative measurement precision of neurological and psychological indicators. At the same time, it allows for a comprehensible interpretation of both types of psychological research: A (non-neuroscientific) psychologist may acknowledge that corroborating evidence can be gained by the neurological approach (the same applies to the cognitive neuroscientist vis-à-vis psychometric data). Identity theory and reflective models view reductive psychological science as an integrated attempt to derive the best measure of the underlying attributes of interest. Such mutually insightful scientific interaction is in line with Heuristic Identity Theory (McCauley & Bechtel, 2001), which argues that simultaneous scientific study of two distinct explanatory levels from an identity theoretical perspective can be mutually beneficial.

Given its attractive theoretical properties, we conjecture that identity theory is implicitly assumed in most cognitive neuroscientific work. However, the conceptual benefits of this application of identity theory come at a price. For example, for both types of indicators to have the same underlying cause, the assumption of unidimensionality must be met. Unidimensionality has testable consequences such as local independence (Hambleton, Swaminathan, & Rogers, 1991) and vanishing tetrads (Bollen & Ting, 1993). If these tetrads are zero (by reasonable approximation), this is an indication that a unidimensional model may be appropriate, or at least, that it cannot be rejected. This suggests that the variability in both psychological and neurological indicators is attributable to a singular underlying cause. The criterion of unidimensionality is strict and certainly need not be satisfied by purported behavioral and neurological measures of a given attribute. Thus, researchers should be clear on whether they believe that their neurological and psychological measurements are truly measuring the same attribute. To summarize, identity theory represents a strict theoretical and statistical position concerning the relationship between the two classes of measurement. It posits that the variability found in both the P- and N-indicators has the same unitary, underlying cause and that the covariance between indicators can thus be fully explained by the underlying cause.

We now consider the integration of neurological and behavioral data from the perspective of supervenience theory. This theory is statistically less restrictive, is conceptually distinct from identity theory, and may provide a more realistic alternative to the stringent requirements of identity theory. In a supervenience conceptualization of psychological processes, the higher order attributes are realized in their neurological properties. This is consistent with a specific implementation of the formative model, called the MIMIC (Multiple Indicators, Multiple Causes) model (Jöreskog & Goldberger, 1975). To illustrate this, a path diagrammatic representation of the MIMIC model of g is displayed in Figure 4. In the MIMIC model, the variability of the determining indicators is a necessary but insufficient condition for variability at the level of the attribute. This is consistent with supervenience theory.

The essential aspect of this model is that there cannot be variation at the latent variable level if there is no variation in the indicators; therefore the theoretical attribute supervenes on its neurological constituents. Conversely, if two people have exactly the same lower order properties, that is, they have the same constellation of relevant neurological activation patterns, they necessarily have the same value on the attribute of interest. The restrictions and characteristics of the strong supervenience thesis and the formative model are identical in this sense. The insufficiency component implies that two people can have different indicator values but the same position at the latent attribute level. Therefore the position on the theoretical attribute is multiply realizable. Accordingly, the mapping of the observations to the theoretical attribute is many-to-one mapping, but no isomorphism, between the indicator values and the attribute value. Moreover, as is generally the case for supervenient properties (Kim, 1992), in the formative model any given position on the theoretical attribute corresponds to a disjunction of lower order properties. For example, a given level of SES may correspond to either having a high salary and poor education, having a low salary and high education, having an average salary and average education, and so on. Thus, the

---

4 A tetrad is the difference of the products of the covariances of four measured indicators.
formative model is an instantiation of the supervenience hypothesis.

A formative approach seems a natural position to take in considering psychological effects of neurological deficits. Consider, for example, Korsakoff’s syndrome. This condition is usually caused by alcohol abuse or malnutrition, which results in neuropsychological symptoms, such as demyelination, neuronal loss, and small-scale hemorrhages (Kopelman, 1995). Psychological manifestations of Korsakoff’s syndrome include impairment in the formation of new memories. In a reflective perspective on Korsakoff syndrome, the behavioral and neurological deficiencies would both be seen as measurement of the presence and severity of the syndrome in a particular patient. This implies a causal direction that runs from the latent variable (a person’s value on a dimension representing the severity of Korsakoff’s syndrome) to the neurological lesions. This seems counterintuitive. A more plausible conceptualization is provided by the formative, or MIMIC, model. Under such a conceptualization, a person’s Korsakoff “score” is determined by a weighted summation of the various lesions, by concurrently measuring and fitting a set of psychologically relevant predictors, such as memory tests. In this case, the lesions are the (partial) causes of Korsakoffs, not vice versa.

The theoretical status of the latent psychological attribute under supervenience theory is distinct from that under identity theory. A researcher who adheres to supervenience theory will represent the latent psychological attribute as being a formative attribute, that is, as being determined by the constellation of neurological indicators. The relative influence of these neurological indicators is estimated on the basis of the predictive ability of the attribute in a network of psychologically relevant predictors.

The supervenience model, as displayed in Figure 4, has two components. The neurological indicators determine the latent psychological attribute. The parameter estimates, or the relative weights of the influence of the neurological measurements (Bollen, 2007), are estimated by predicting a psychologically relevant set of attributes or behaviors. The reflective component of a supervenience model is often required to be unidimensional. However, the formative part of the model is not so constrained: The indicators may even be uncorrelated (Bollen, 1984; Curtis & Jackson, 1962). This model is therefore less restrictive than a reflective, identity theoretical model. To summarize: An individual’s position on a formative latent attribute, under the theory of supervenience, may be estimated by fitting the model to a set of behaviorally predictive psychological measurements. The identity of the attribute is determined by the neurological attributes included in the model that specifies the strength and direction of the neurological indicators. These indicators are assumed to determine variability in the latent attribute, which in turn determines variability at the psychological process.

The different empirical planes of the N indicators and the P indicators in a supervenience conceptualization, as opposed to identity theory, are important to neuroscience. The psychological indicators are scores derived from measurement instruments that are used in the model specification. The parameter estimates, which relate variability in the latent attribute to variability on the N indicators, depend on which P-indicators are chosen in the model. However, it is possible that the same set of N-indicators will fit models with different sets of P-indicators. Thus, the same N-indicators may realize different latent variables, as specified by different sets of P-indicators. This is a significant difference with the identity model, in which this is impossible.

This is important because it establishes that, given supervenience, the identification of attributes, even if they are neurologically grounded, depends on the psychological, not the neurological, part of the model. This is consonant with the finding that certain neural structures are “implicated” in a wide range of different psychological concepts and processes. For
example, the dorsolateral prefrontal cortex has been found to be differentially active in processes as psychologically diverse as response selection (Holland, Rushworth, Passingham, Jahanshahi, & Rothwell, 2001), pain modulation (Lorenz, Minoshima, & Casey, 2003), components of working memory (Ranganath, Johnson, & D’Esposito, 2003), voluntary willed action (Frith, Friston, Liddle, & Frackowiak, 1991), response inhibition (Ridderinkhof, Wildenberg, Segalowitz, & Carter, 2004), mastication (Takahashi, Miyamoto, Terao, & Yokoyama, 2007), schizophrenia (Weinberger, Berman, & Zec, 1986), and intelligence (Jung & Haier, 2007). For an equally heterogeneous assessment of the functions of the anterior cingulate, see Vogt, Finch, and Olson (1992) or Devinsky, Morrell, and Vogt (1995). This functional heterogeneity should not be construed as a failure of cognitive neuroscience but rather as an inherent property of brain function and organization. The point is that if certain cortical areas are associated with different cognitive functions, then it is unlikely that fMRI activity in such an area can be considered, for example, a “measurement of” working memory, as the assumption of unidimensionality will probably not be met.

We cannot think of an a priori reason to prefer either the identity or the supervenience model. Instead, we think that appropriateness of either model will depend on the attribute that is being studied and on theoretical considerations concerning that attribute. However, we note that precisely these theoretical considerations may be of great conceptual assistance to reductive psychological science, as they force researchers to consider the status of the attribute they are interested in and the most appropriate manner to study it. Our argument here is that such choices are not esoteric statistical considerations: They concern unavoidable assumptions implicit in any type of reductive research. The goal of this approach is twofold: Positions from philosophy of mind can be made empirical,5 and empirical neuroscience is provided with a method to get a grip on some of the more nebulous metaphors concerning the relation between psychological and neurological properties.

Doing so may yield several benefits, most notably the avoidance of ambiguous interpretations that may otherwise arise. If the issues previously mentioned are not addressed explicitly, the questions being studied and the interpretations of the data may suffer. Consider, for instance, Jung and Haier (2007), who raised the question, “Where in the brain is intelligence?” (p. 135). Jung and Haier examined 37 methodologically heterogeneous studies that reported correlations between various measures of intelligence and the brain. Their model, called the Parieto-Frontal Integration Theory model, is built on the basis of what are, in the words of Norgate and Richardson, “correlations between those correlations” (2007, p. 162) and describes what happens when an individual is involved in intelligent behavior (p. 138). Although the effort of combining insights from various studies is commendable, the conceptual ground for interpreting the correlations between intelligence and brain measures in this review is at times unclear, and findings are therefore hard to interpret.

First, the question asked by Jung and Haier implies the possibility of the localization of intelligence. However, as intelligence is an interindividual construct, this is akin to the question, “Where in the body is tallness?”—a confusing question at best. Tallness is a property of the body; it does not reside in it. Similarly, intelligence is a property of the cognitive system and does not reside in a particular part of the brain. Second, despite being based on interindividual differences, the Parieto-Frontal Integration Theory model is in essence an intraindividual model of intelligent behavior. However, as Borsboom, Mellenbergh, and van Heerden (2003) and Molenaar (2004) showed, these two domains are quite distinct: Results at the population level are not necessarily informative about the individuals that make up that population. This discussion illustrates that analyzing and interpreting relations between neurological and behavioral measurements can benefit from a sound conceptual basis. To illustrate how psychometric models may be able to provide more insight, we examine the application of the two previously discussed models to two empirical examples, focusing on neurological measurements with respect to personality characteristics and general intelligence.

**Intelligence and Brain Volume**

To illustrate the issues we just discussed, we return to the question we posed in the introduction: Can a measurement of the volume of a person’s brain be considered a measurement of their intelligence? One of the more robust findings in the literature relating intelligence to physiological characteristics is the relationship between skull (or more recently brain) volume and estimates of general intelligence. Based on a meta-analysis, this correlation has been estimated at .33 (McDaniel, 2005). Given this relatively solid statistical association, can we consider measurements of brain volume to be measurements of intelligence, and therefore to conform to the identity theoretical perspective? That is, do measurements of brain properties and intellectual ability together fit a unidimensional reflective model?

---

5 The idea that epistemological speculation can gradually be replaced by empirical science, sometimes termed naturalism or naturalized epistemology, has a long history. See, for instance, Quine (1969) for a philosophical motivation.
Methods

To examine this question empirically, we consider behavioral measures (intelligence tests) and physiological (brain mass volume) measures. The sample consisted of physiological and behavioral data acquired from 80 healthy participants (mean age 21.1 years, SD = 2.55, 29 male, 51 female). The measures of intelligence are four domain scores of the commonly implemented Wechsler (2005) Adult Intelligence Scale (WAIS–III). The domain score subscales used were Verbal Comprehension ($M = 117.16, SD = 9.78$), Perceptual Reasoning ($M = 112.10, SD = 11.31$), Working Memory ($M = 111.32, SD = 13.11$), and Processing Speed ($M = 116.38, SD = 14.80$). In addition to the behavioral measurements, all participants were scanned to estimate white matter, gray matter density, and cerebrospinal fluid volume. Details of the scanning procedure and preprocessing steps are described in the appendix. To determine model fit, we examined the chi-square test of model fit, the RMSEA (cutoff value = 0.05), the CFI (cutoff value = 0.95) Akaika Information Criterion (AIC; Akaika, 1974), and the Bayesian Information Criterion (BIC; Schwarz, 1978).

For both models, the first reflective parameter was scaled to 1 to identify the reflective parameters. For discussions on the relative merits of these indicators, see Hu and Bentler (1999), or Schermelleh-Engel et al. (2003).

We consider this experimental setup from the perspective of the two models that we previously discussed. In fitting both models we use the same data but impose distinct constraints consistent with the two models. In both models we view “intelligence” as an attribute that can be studied by psychological and physiological measurements, even though it cannot be observed directly. From the perspective of the reflective model, we consider both methods of measurement (i.e., voxel-based morphometry [VBM] and the WAIS) as measurements of intelligence, in the same way that an electrical and a mercury thermometer may both measure temperature. This conceptualization has been represented previously in Figure 3. For this specific implementation, we would have four psychological measures and three neurological measures measuring the same property ($g$).

Conversely, one may view the neurological measurements as determining the latent psychological attribute. For instance, we may conjecture that the brain volume determines the level or degree of intelligence, in the same way that we know that physiological damage can affect personality. In this case, we consider the MIMIC model to be appropriate. This is the model previously represented in Figure 4, in the MIMIC model of general intelligence and brain characteristics, the neurological indicators determine the value of the latent attribute (i.e., the $g$ score). This in turn can be seen as the underlying cause of the variability of the scores at the WAIS level.

Model Fit Comparison

We used Mplus (Muthén & Muthén, 1998–2007) to fit the reflective and the formative (MIMIC) models for these seven indicators using maximum likelihood estimation. First, we examined the simple reflective model, in line with identity theory. The model was rejected by the chi-square test of model fit, $\chi^2(14, N = 80) = 51.6, p < .01$. The other fit indices corroborate this poor fit (CFI = 0.88, RMSEA = 0.18, AIC = 3706.39, BIC = 3739.74). For this data set therefore, Identity Theory is rejected, and we cannot consider measurements of brain volume to be measurements of intelligence. Next, we considered the MIMIC model, in line with supervenience theory. This model fits the data well. The model was not rejected by the chi-square test of model fit, $\chi^2(11, N = 80) = 11.20, p > .4$. Other fit indices supported the good fit of the model (CFI = .996, RMSEA = 0.015, AIC = 3659.994, BIC = 3686.196). Table 1 shows the parameter estimates for both models, which quantify the relative strength of the relationship between the indicators and the latent attribute “$g$.”

For this data set therefore, a reflective model (identity theory) does not fit the data. The MIMIC (supervenience) model, on the other hand, fits the data quite well and explains .25 of the variance in general intelligence, in line with previous analyses, clearly favoring this model for this data set. However, the distinction in model fit will not be as clear-cut for all psychological constructs. Next, we examine a data set where the distinction is less pronounced.

Personality and the Brain

Another type of construct traditionally of interest for scientific psychology is that of personality. One

### Table 1. Parameter Estimates for Reflective and Formative (MIMIC) Models of Intelligence.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Reflective Model</th>
<th>MIMIC Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAIS1</td>
<td>0.273</td>
<td>0.601</td>
</tr>
<tr>
<td>WAIS2</td>
<td>0.236</td>
<td>0.714</td>
</tr>
<tr>
<td>WAIS3</td>
<td>0.301</td>
<td>0.534</td>
</tr>
<tr>
<td>WAIS4</td>
<td>0.246</td>
<td>0.5</td>
</tr>
<tr>
<td>Gray matter volume</td>
<td>0.983</td>
<td>0.883</td>
</tr>
<tr>
<td>White matter volume</td>
<td>0.967</td>
<td>−0.714</td>
</tr>
<tr>
<td>CSF</td>
<td>0.752</td>
<td>0.304</td>
</tr>
</tbody>
</table>

Note. MIMIC = Multiple Indicators, Multiple Causes; WAIS = Wechsler Adult Intelligence Scale; CSF = Cerebrospinal Fluid. *Standardized factor loading.
of the more famous models is the Big Five model of personality (McCrae & John, 1992), which describes variation in personality traits along five dimensions (Extraversion, Neuroticism, Conscientiousness, Openness, and Agreeableness). Certain aspects of personality have been shown to correlate with differential brain activity and physiology (DeYoung & Gray, 2009; Wright et al., 2006). In fact, one of psychology’s most famous case studies, that is, the case of Phineas Gage, suggests that brain physiology may be of significance to researchers of personality (Damasio, Grabowski, Frank, Galaburda, & Damasio, 1994). We examine the conceptual and statistical relationship between psychological data on a common personality subscale, Conscientiousness, on one hand, and a physiological measurement—in this case, gray matter density—on the other hand.

Methods

In this study, physiological and behavioral data were acquired from 110 healthy participants (age M = 21.4, SD = 2.4, 27 male). The participants were tested on the abbreviated personality questionnaire NEO-PI (McCrae & Costa, 2004). This personality questionnaire comprises 60 items, with 12 items for every Big Five personality dimension (i.e., Extraversion, Neuroticism, Conscientiousness, Openness, and Agreeableness). For the purpose of this illustration we focus on one subscale, Conscientiousness. In addition, we obtained of each subject two 3DT1 scans to study VBM. VBM is a voxel-wise comparison technique that uses high-resolution structural scans to estimate gray matter density values at the voxel level (Ashburner & Friston, 2000, 2001). Eight participants were excluded due to recording problems or the lack of a second scan, leaving 102 participants for subsequent analysis. We provide further preprocessing and scanning details in the appendix. As with the general intelligence data, we fit two models: a reflective model in line with identity theory, and a MIMIC model in line with supervenience theory.

Model Fit Comparison

We used Mplus (Muthén & Muthén, 1998–2007) to fit the reflective and the formative (MIMIC) models using maximum likelihood estimation. Using an iterative procedure that excluded parameters if model fit improved significantly by their removal, the final models included four brain regions (Left Supramarginal Gyrus, Right Middle Frontal Gyrus, Left Cerebellum, Right Cerebellum) and 11 of the original 12 conscientiousness questions.

First, we considered the reflective model, in line with identity theory. The reflective model was rejected by the chi-square test, $\chi^2(90, N = 105) = 120.49, p < .05$. The other fit indices corroborated the poorer fit of the reflective model (RMSEA = 0.06, CFI = .84, AIC = 2129.21, BIC = 2207.96). Second, we considered the MIMIC model. This was not rejected by the chi-square test of model fit, $\chi^2(84, N = 105) = 100.65, p > .10$. The other fit indices suggest reasonable fit (CFI = 0.91), RMSEA (0.04), AIC (2101.37), and BIC (2169.62). Table 2 shows the parameter estimates for both models, which quantify the relative strength of the relationship between the indicators and the latent attribute “conscientiousness.”

Because these two models are by their nature not nested, a chi-square test to compare them directly is not possible (Vuong & Wang, 1993). However, the formative (MIMIC) model shows better fit across the board than the unidimensional reflective model, with all fit indices outperforming those of the reflective model. Overall then, this suggests that the formative model provides a better fit to the data than the reflective model. The present study thus provides some support for a supervenience interpretation of the relation between neurological and psychological variables with respect to conscientiousness.

Implications

As we show earlier, it is possible to fit such models to conventional neuroimaging data. There are several important aspects of the two illustrations. First, the reflective, identity theoretical model

Table 2. Standardized Parameter Estimates for Reflective and Formative (MIMIC) Models of Conscientiousness.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Reflective Model$^a$</th>
<th>Formative (MIMIC) Model$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.361</td>
<td>0.359</td>
</tr>
<tr>
<td>C2</td>
<td>0.231</td>
<td>0.23</td>
</tr>
<tr>
<td>C3</td>
<td>0.207</td>
<td>0.208</td>
</tr>
<tr>
<td>C4</td>
<td>0.336</td>
<td>0.337</td>
</tr>
<tr>
<td>C5</td>
<td>0.736</td>
<td>0.738</td>
</tr>
<tr>
<td>C6</td>
<td>0.397</td>
<td>0.398</td>
</tr>
<tr>
<td>C7</td>
<td>-0.204</td>
<td>-0.203</td>
</tr>
<tr>
<td>C8</td>
<td>0.313</td>
<td>0.315</td>
</tr>
<tr>
<td>C9</td>
<td>0.226</td>
<td>0.225</td>
</tr>
<tr>
<td>C10</td>
<td>0.802</td>
<td>0.797</td>
</tr>
<tr>
<td>C11</td>
<td>0.73</td>
<td>0.731</td>
</tr>
<tr>
<td>Left supramarginal gyrus</td>
<td>-0.3</td>
<td>-0.303</td>
</tr>
<tr>
<td>Right middle frontal gyrus</td>
<td>0.29</td>
<td>0.311</td>
</tr>
<tr>
<td>Left cerebellum</td>
<td>0.062</td>
<td>0.124</td>
</tr>
<tr>
<td>Right cerebellum</td>
<td>-0.001</td>
<td>-0.062</td>
</tr>
</tbody>
</table>

Note. MIMIC = Multiple Indicators, Multiple Causes.

$^a$Standardized factor loading.

---

6Eighty of the participants in the personality data set were also analyzed in the intelligence data set, albeit on different behavioral and neurological measurements.
was rejected for both data sets. Despite the adequate sample size and neurological variables known to correlate with the respective constructs, we cannot consider such measurements, for these data sets, to be measurements of the psychological constructs of interest. This points to an interesting conclusion that follows from the identity hypothesis: Researchers who view brain measurements as measurements of a latent psychological attribute (which may be plausible), must realize that this accords to brain measurements the same status as psychological measurements. Consequently, the brain measures may be rejected for the same reasons that poorly performing items in a questionnaire are rejected. This illustration shows what is required of neurological measurements if they are to figure as “measurements of” attributes in the same way that psychological measurements do. However, although the reflective model did not fit for the two examples we examined, this does not imply it will not fit for any data set.

First, the strength and nature of the relationship between psychological construct and neurological properties will vary depending on the construct, as it does in our two data sets. Given the variability of the psychological constructs that figure in scientific psychology, from early visual perception to complex dispositional constructs, it seems likely that the strength and nature of the relationship between neurological and behavioral measurements will be also different for such radically different behavioral phenomena. Second, we think it more likely that more restrictive models, such as the reflective model, will fit for more “basic” and less variable, psychological constructs and processes. For instance, whereas personality dimensions are at least partly culturally determined, other processes such as retinotopic mapping of early visual processing (mentioned in the introduction), depth perception, and arousal may be easier to identify with unique, unidimensional neurological signatures. Such lower, more basic, less culturally dependent constructs, which display less variance across people, may be good candidates for identity theoretical models, although at this point this is largely speculation.

For both data sets, especially the brain volume data, the MIMIC model fit the data quite well. This implies that in these data sets, it is sensible to conclude that the neurological measurements statistically determine the variability in the psychological construct. Most important, the current findings show us that the relationship, especially for more complex psychological constructs such as intelligence and personality dimensions, is not likely to be simple. For this reason above all, we should be closely examining the nature of this relationship and try to gain more insight by modeling hypotheses explicitly.

Summary

The results of our model fitting and speculations about other constructs bring to light an important aspect of the present reformulation of the reduction problem as a measurement problem: At the outset of any investigation, we should be impartial with respect to the status and quality of psychological and neuroscientific assessments as measures. For example, aspects of personality have been called “biologically based tendencies” (McCrae et al., 2000, p. 173). It remains to be seen whether certain empirical measurements behave in a way that allows for such an interpretation. Despite the popular view of neuroscientific measures as being “exact” or “hard,” at least for this data set our analysis suggests that psychological measures may outperform neuroscientific measures. Insofar as such measures are interpreted as relevant to psychological attributes or processes, they should be evaluated on precisely the same basis as any other measure. This basis is psychometric in character. There is no way out of this issue, unless perhaps one comes up with an alternative to psychometrics, that is, a practically workable theory of measurement that rests on a different basis. To the best of our knowledge, such a theory does not currently exist.

The aforementioned empirical illustration serves as a proof of principle, in that it demonstrates that conceptual positions about the relationship between two classes of data can be constructed as statistical models and empirically tested. In this manner conceptual ideas about the relation between two levels of measurement can be translated into falsifiable models and allow for theoretical interpretations of empirical data that can go beyond the simple observation that two measures are associated. In the next sections, we discuss certain practical issues concerning SEM models and the possibility of more exotic extensions.

Applying SEM in Practice

Although SEM models generally require larger sample sizes than more conventional analysis methods, this increase is by no means prohibitive. Sample sizes for SEM models are, as with other statistical analyses, related to model complexity. The models we discuss here are relatively simple, and sample sizes required are well within the reach of practicing neuroscientists. For instance, Marsh and Hau (1999) showed that for models with 6 to 12 indicators per factor (as is the case for both our data sets), sample sizes of 50 may be adequate. Bentler (1995) recommended a total sample size of at least 5 per free parameter, again within the limits of our empirical illustrations (cf. Schermelleh-Engel, Moosbrugger, & Müller, 2003, for an discussion on this topic). Although it is true that more simple or
conventional data-analytic procedures such as analyses of variance and correlation tests yield results for very small sample sizes, over the long run such analyses may represent pyrrhic victories over issues of inferential validity and replication.

Luckily, there are signs that the field of neuroimaging has increasingly moved toward sample sizes that are more than adequate for treatment with SEM. For instance, a quick inspection of the first eight empirical articles in a recent issue of the journal Neuroimage (Volume 51, Issue 1, the Anatomy and Physiology section) that focus on structural anatomy (such as we examine in our article) reveal sample sizes of 90, 55, 319, 185, 70, 40, 45, and 280 respectively, all of which would be amenable to SEM approaches given the aforementioned guidelines. At the same time, it is certainly true that lower sample sizes are a common occurrence in neuroimaging, especially in functional neuroimaging studies. However, to deal with the inherent complexity of the relationship between the brain and psychological constructs, more complex models, which require greater sample sizes, will need to be developed. The move toward larger, more versatile data sets may be part of a broader development in the field of cognitive neuroscience, taking inspiration from how neighboring fields deal with similar problems.

In the field of quantitative genetics, issues of replicability, power, and interpretation have led to the realization that larger sample sizes are not a luxury but a necessity. This realization has lead to large collaborative projects such as the EAGLE and the GENEQOL consortium. Such large-scale collaborative efforts combine the knowledge, resources, and methodology from various research groups; can lead to increased collaboration and understanding; and therefore benefit the scientific community as a whole. It is such collaborations that would make the implementation of more insightful models possible and, in our view, would benefit the field as a whole. An additional advantage of the use of SEM in the context of collaborative projects is that there exist a statistically and theoretically sound way to deal with group differences (i.e., measurement invariance; Meredith, 1993). A similar development in the fields of cognitive, affective and social neuroscience would be much welcomed (the Brainmap www.brainmap.org, represents a considerable step in this direction).

To summarize, the sample sizes required to test conceptually guided SEM models are well within the reach of current empirical practice. To the extent that such data sets are not yet widely available, larger collaborations are desirable. Such collaborations are especially important if we want to tackle some of the most elusive and vexing phenomena: dynamic, reciprocal changes over time. In the next section, we show how philosophy of mind and extensions of basic SEM models may help to get a grasp on such phenomena.

**Top-Down Influences and Temporal Dynamics**

The models that we have discussed represent two core philosophical positions, which have well-defined SEM counterparts. So far, we have focused on the most conventional method of analysis: the analysis of interindividual differences in cross-sectional data. This method is dominant in contemporary psychological science. Although this method provided the basis of our proposed structuring of the relationship between behavioral and neurological data, other methodological approaches are possible. In fact, there are aspects of psychological and neurological phenomena that may be better studied by alternative means. In this section, we discuss some challenging problems for conceptual and statistical models in cognitive neuroscience. These concern dynamic, reciprocal changes of behavior, and brain structure and function through time. We note that SEM offers various possibilities to address these problems.

Conventional thought concerning the relationship between lower and higher order properties tends to consider (changes in) neurological properties as the source or cause of observable difference at the psychological/behavioral level. For instance, evidence shows that certain drugs influence cognitive abilities (Maylor & Rabbitt, 1993), that trauma may influence complex psychological traits such as personality (Damasio et al., 1994), and that in Alzheimer’s patients amyloid peptide levels (constituents of amyloid plaques) and cortical activity are affected *prior* to observable cognitive symptoms (Buckner et al., 2005; Moonis et al., 2005). These findings all suggest that changes in cortical structure or functioning can, and do, affect psychological performance and functioning. However, there is also ample evidence for the reverse causal path. For instance, Maguire et al. (2000) showed that London taxi drivers, following intensive training to learn the streets of London to the mandatory level of competence, showed structural changes to the hippocampus, and these changes were greater for taxi drivers who had served for a longer time. Also, intense juggling practice has been shown to affect both gray matter density (Draganski et al., 2004) and white matter integrity (Scholz, Klein, Behrens, & Johansen-Berg, 2009). These findings suggest that persistent behavior may affect neurological structure and functioning. Finally, processes may even be *reciprocal* in nature, that is, simultaneous influences both from neurophysiological properties to behavior and vice versa. For instance, take the influence of hormone levels on psychology and behavior.

---

Testosterone, when injected, can directly influence dominant or aggressive behavior and is found to correlate positively with such behaviors (Mazur & Booth, 1998). However, Mazur and Booth illustrated that causal relationship may also be reversed: Certain psychological behaviors may themselves lead to an increase in testosterone, and elevated testosterone levels affect behavior.

The aforementioned examples suggest that influences may run both from cognitive/psychological processes to neurological changes and vice versa. Modeling such dynamic interactions over time is a challenging problem, both conceptually and statistically. Philosophers have discussed such complex, dynamic systems in various terms, such as emergence, dynamic systems, and top-down causation. Emergence has a long philosophical tradition, going back to Mill and Broad in the late 19th and early 20th century (for an overview, see Kim, 1999). Recently, emergence and dynamic systems have enjoyed renewed interest as possible models for dynamic, neurocognitive changes through time. For instance, Jost, Bertschinger and Olbrich (2010) discussed the philosophical construct of emergence, and the description of a neurosystem as a (nonlinear) dynamical reciprocal system. Similarly, Walmsley (2010) examined the concept of emergence, its relevance for complex systems, and possible manners in which lawlike properties may emerge at higher (psychological) levels. Craver and Bechtel (2007), on the other hand, focused on the concept of downward causation and how this may be reconciled philosophically. They concluded, “When interlevel causes can be translated into mechanistically mediated effects, the posited relationship is intelligible and should raise no special philosophical objections” (p. 547). Furthermore, they stated, “There is a different sense in which a cause can be said to be at the top (or bottom) and a different sense in which its influence is propagated downward (or upward)” (p. 548).

Here we attempt to structure the distinction between such interlevel effects. Certain SEM models offer the means to study such complex, interactive processes empirically. The origins of these models can be traced to the 1930s and 40s (e.g., Bartlett, 1946), but specific implementations in the behavioral/psychological sciences are relatively new. For instance, Hamaker et al. (2007) described a time series model in which two latent variables and their respective influences were modeled over time. We discuss how such a model may be used to model the time course of complex phenomena, such as previously discussed. We distinguish two “types” of situations: those in which psychological behavior affects, or at least precedes, variation in neurocognitive changes through time. For instance, Jost, Bertschinger and Olbrich (2010) discussed the philosophical construct of emergence, and the description of a neurosystem as a (nonlinear) dynamical reciprocal system. Similarly, Walmsley (2010) examined the concept of emergence, its relevance for complex systems, and possible manners in which lawlike properties may emerge at higher (psychological) levels. Craver and Bechtel (2007), on the other hand, focused on the concept of downward causation and how this may be reconciled philosophically. They concluded, “When interlevel causes can be translated into mechanistically mediated effects, the posited relationship is intelligible and should raise no special philosophical objections” (p. 547). Furthermore, they stated, “There is a different sense in which a cause can be said to be at the top (or bottom) and a different sense in which its influence is propagated downward (or upward)” (p. 548).

Here we attempt to structure the distinction between such interlevel effects. Certain SEM models offer the means to study such complex, interactive processes empirically. The origins of these models can be traced to the 1930s and 40s (e.g., Bartlett, 1946), but specific implementations in the behavioral/psychological sciences are relatively new. For instance, Hamaker et al. (2007) described a time series model in which two latent variables and their respective influences were modeled over time. We discuss how such a model may be used to model the time course of complex phenomena, such as previously discussed. We distinguish two “types” of situations: those in which psychological behavior affects, or at least precedes, variation in neurocognitive changes through time. For instance, Jost, Bertschinger and Olbrich (2010) discussed the philosophical construct of emergence, and the description of a neurosystem as a (nonlinear) dynamical reciprocal system. Similarly, Walmsley (2010) examined the concept of emergence, its relevance for complex systems, and possible manners in which lawlike properties may emerge at higher (psychological) levels. Craver and Bechtel (2007), on the other hand, focused on the concept of downward causation and how this may be reconciled philosophically. They concluded, “When interlevel causes can be translated into mechanistically mediated effects, the posited relationship is intelligible and should raise no special philosophical objections” (p. 547). Furthermore, they stated, “There is a different sense in which a cause can be said to be at the top (or bottom) and a different sense in which its influence is propagated downward (or upward)” (p. 548).
can be extended to include additional psychological or neurological latent variables, with varying numbers of indicators.

If we measure the indicators of two (or more) latent variables in a given person repeatedly over time, we can relate the indicators to the latent variables, and we can model the time series of the latent variables. As previously mentioned, we would like to differentiate between two scenarios: Variability in the psychological construct precedes neurological variability, and vice versa. If we were to assess a sample of people over time, either improving in juggling or deteriorating in cognitive performance, we could model this process by means of psychometric models such as the Integrated State-Trait model, described by Hamaker et al. (2007).

To ease our presentation, we have included in Figure 5 only the most relevant parameters. The model parameter \( \phi_P \) in Figure 5 represents the influence of the psychological latent variable at a given time point on the neurological latent variable at the next time point. This parameter should deviate from zero if changes in psychological abilities or behavior (e.g., practicing juggling) affect changes in neurological substrate (e.g., gray matter density in a motor cortex region). Conversely, the parameter \( \phi_N \) reflects the influence of the neurological latent variable on the psychological latent variable. This parameter should deviate from zero if neural changes (e.g., amyloid peptide levels) affect changes at the psychological level (e.g., psychological performance).

So given appropriate time series measurements, we can test the hypothesis that the variation at the neurological precedes variation at the psychological level, or vice versa. In doing so, it is possible to empirically distinguish cases where influence should best be represented as “bottom up” or “top down.” Note that in this model we purposely estimate the relationship and not assume it a priori: For this reason, a MIMIC model is not appropriate. Rather, we want to explore the time course of possibly reciprocal influences to gain insight into the nature of the underlying processes. Despite the complex nature of such dynamic processes, SEM models allow researchers, at least in principle, to get a grip on the structure of the development over time and reciprocal interactions.

Discussion

The scientific future of reductive cognitive neuroscience research rests both on advances in brain scanning technology and on the development of a comprehensive conceptual framework to link psychological-behavioral measures and neurological measures. To do so, a careful consideration of the status of neurological indicators in studies that measure both behavioral and neurological variables is required. We have shown a road forward in attacking this problem by demonstrating that at least two theoretical stances on the reduction problem can be translated into well-understood formal psychometric models. To our knowledge, this is the first demonstration of how theoretical positions drawn from analytic philosophy can be translated to empirically testable models. Notably, our demonstration did not involve any rocket science; it merely used standard statistical models incorporated in widely available software packages. In this regard, the suggested models are ready for use, and there is little that stops the motivated researcher from utilizing their benefits.

Several of the issues raised by other authors we discussed in the introduction can be ameliorated, if not solved, by our proposed framework. First, by explicitly framing the connection between neurological and behavioral variables as a measurement theoretical relationship, the mereological fallacy can be largely avoided. The models proposed here do not make claims about certain psychological processes being “in” or “performed by” a certain brain region any more than the answers of the questionnaire are the locus of personality traits. In fact, we would argue that the current perspective offers a way to discuss brain–behavior relationships in a meaningful manner without running the risk of making mereological or seemingly neophenological claims.

Furthermore, the correct application of SEM models (greatly) diminishes the problem of nonindependence raised by Vul et al. (2009). The two-step procedure described in Vul et al. can be largely avoided by properly implementing formal measurement models. As the voxels are not treated as a large sets of independent statistical tests, but specified as part of a measurement model that implies certain covariance patterns, the multiple comparison issue is much less of a problem given the considerable size of data sets within neuroscience.

Finally, our approach can accommodate some of the ideas put forth by Barrett (2009). She argued that certain psychological processes may be more appropriately seen as a “mix” (or “recipe”) of several classes or types of brain activity. That is, the categorical distinctions we make at the psychological level, for example, between “thinking,” “perceiving,” and “remembering,” will probably not be found as categorically distinct processes or properties of the brain. For that reason, when studying neurological properties in relation to certain psychological properties, it may be more natural to think that different combinations of distributed activity in certain regions or systems can together be taken to represent distinct brain processes. According to Barrett, this more naturally accommodates the structure of the brain than old-fashioned perspectives such as positing a “perception” region in contrast to a “memory” region. Within the framework currently proposed, such hypotheses (that categorically distinct psychological concepts may be best seen as complex combinations
of more basic processes) may be tested. This is best in line with the supervenience/MIMIC model. Given different psychological predictors (i.e., whether the reflective part of the model consists of personality items, Raven’s matrices, etc.), one would expect to find different parameter estimates of the neurological measurements. The different weighting of neurological indicators is conceptually similar to the recipe metaphor proposed in Barrett. This underscores the flexibility of the current approach of implementing measurement models to test substantive hypotheses about brain–behavior relationships.

We hope that this article has served to convince the reader that the infamous reduction problem is at least partly a measurement problem. More specifically, one cannot hope to make true advances in solving the reduction problem without solving the associated measurement problems in parallel. This, we think, has substantial consequences for how we should evaluate reductionist claims, as well as what we can expect from reductionist research strategies. There are several reasons to pursue such a strategy.

There is a tendency, both in science and society, to view neuroscience as an exact area of research—closely related to physics, chemistry, and biology—while viewing psychology as a “soft” discipline (cf. Racine, Bar-Ilan, & Illes, 2005). However, the exact sciences are not exact because they use machines rather than questionnaires, but because they have successfully formalized theories. Such formalization is currently lacking at the interface of neuroscience and psychology. Thus, insofar as neuroscience has moved into the field of psychology, it has yet to earn the predicate of being a “hard” science. Escape from this situation can only be realized by formalizing theories into mathematical models, which are likely to be statistical in nature, and insofar as these models concern measurement problems, they will likely be psychometric ones. For models to function properly, there should be no psychometric prejudice as to the quality of the measurements. From a measurement perspective, neurological measurements do not have a privileged position over conventional psychological measurements.

To the researcher with expertise in the intricacies of psychometric modeling, the example illustration in this article may be viewed as quite optimistic, and such an evaluation would not be entirely off the mark. For even though we think it is evident that psychometrics has much to offer to neuroscience, it should be noted that psychometric modeling can be quite complicated. For instance, as discussed before, successful modeling generally requires (slightly) larger sample sizes or extensive time series, attention to possible problems involving model identification and model equivalence (e.g., see Raykov & Penev, 1999), goodness-of-fit, and other general issues common to statistical modeling. However, we think that such issues, in general, do not pose greater problems for SEM models than for other techniques and should not detract from substantively guided model implementation. There really is no way around these problems; in particular, these issues will not be resolved by being ignored and by proceeding as if one did not have a measurement problem to solve.

The models we have discussed in the present work are illustrative of how clean and simple identity and supervenience theories really are. As a result, it is likely that the models that we applied to personality measures may be too simplistic. This, however, is a benefit rather than a shortcoming of the psychometric representation of reductionist theories: A psychometric representation makes the hypotheses proposed transparent and subject to informed criticism, and it does this to a degree that no verbal description could match. Moreover, rejecting these models brings with it the task of inventing better ones. And this, we think, is precisely the road to progress. In addition, it is likely that alternative models will lead to alternative philosophical views on the relation between psychology and neuroscience. We have provided a proof of principle by fitting two models with a single latent variable measured at the interindividual level. However, our approach is certainly not limited to such a design; one of the great benefits of SEM models is their flexibility. Most psychological processes involve a complex interplay of more than one attribute. For example, complex cognitive processes such as problem solving almost surely involve the interaction of separate subsystems such as working memory, attention, and intelligence. SEM models can be extended to include multiple latent variables, thereby testing hypotheses about the interactive, inhibitory, or excitatory activity of several latent variables of psychological attributes within a measurement model. The theoretical approach discussed here is especially suited for the implementation of flexible models that may address a range of questions of substantive interest to both cognitive neuroscientists and philosophers.

Finally, there are at least two types of homogeneity within cognitive neuroscience that are often assumed rather than tested. For instance, one of the vexing and largely neglected issues within psychological science is the distinction between inter- and intraindividual explanation. This means that a result found at the group level is often taken to apply to the individual, despite the fact that this may not be true and is rarely tested (cf. Borsboom et al., 2003; Molenaar, 2004; Molenaar, Huizenga, & Nesselroade, 2002). This issue holds as much for cognitive neuroscience as it does for conventional psychological science. As described previously, the current approach can be extended to explicitly test the structure of intraindividual activation. For example, the time series model by Hamaker et al. (2007)
showed how an intraindividual process can be modeled with repeated measurements of the same latent variable. This can be done in much the same way within the current framework by including dynamic intraindividual measurements such as EEG or fMRI. In this manner, one can study the extent to which a latent variable at the interindividual level is representative for the individuals that make up a population, by assessing the homogeneity of the latent variable at both levels. Interindividual variability is commonly treated as measurement error, but by explicitly testing the tenability of this assumption, a more fine-grained understanding of psychological attributes may be possible. In fact, the extent to which this holds for certain psychological attributes but not for others is likely to yield valuable insights. Another largely neglected but potentially insightful area of cognitive neuroscience is the question of homogeneity across subpopulations. Within the current framework, it is relatively easy to test whether a latent variable representation of, say, working memory differs across age groups, gender, or other subpopulations (e.g., see Meredith, 1993). For example, Henrich, Heine, and Norenzayan (2010) examined to what extent it is possible to generalize from the most commonly studied psychological subpopulation, namely, young, White, highly educated people from industrialized nations, to other cultures and demographics. They showed that, even for the most “basic” of cognitive phenomena such as the Mueller–Lyer illusion, such untested generalization is often unjustified. The assumption of generalization and homogeneity is, arguably, even more omnipresent within cognitive neuroscience than in conventional psychology. We would venture that the extent to which neuroscientific findings generalize across populations and cultures is an open empirical question and that its premature acceptance may close off a considerable amount of potentially insightful empirical investigations.

This article has served to illustrate both the necessity and the potential for conceptual and empirical progress that may be achieved by considering an integrated psychometric perspective on reductive cognitive neuroscience. We have offered the conceptual and technical tools to do so, and we hope that our efforts will be built on by others. The relationship between mind and body has fascinated generations of philosophers and scientists, and it deserves closer methodological and psychometric scrutiny than it has so far enjoyed. If theories developed in the philosophy of mind are to escape from their current state of splendid metaphysical isolation, it is essential to translate these positions to empirical predictions. With recent advances in neuroscientific and psychometric techniques and methods, we finally have the opportunity to empirically address questions that were once restricted to the realm of speculative metaphysics. It would be a waste to forgo such opportunities.

Acknowledgments

This article is original, and the material has not been published elsewhere. The participants were tested in accordance with the ethical guidelines of the American Psychological Association, and this research was approved by the University of Amsterdam Ethical Committee.

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


Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., & Hawkes, D. J. (1999). Nonrigid registration using free-
APPENDIX

Imaging and Preprocessing

Participants were scanned on a 3-T Philips Intera scanner, and all data were analyzed using FSL (Smith et al., 2004), Matlab (Mathworks Inc.), and Mplus (Muthén & Muthén, 1998–2007). A structural MRI scan of each participant was acquired using a T1-weighted 3D sequence (Turbo Field Echo, TE 4.6 ms, TR 9.6 ms, FA 8°, 182 sagittal slices of 1.2 mm, FOV 250² mm, reconstruction matrix 256²).

For the study on intelligence we first extracted the brains from the structural images (Smith, 2002) and subsequently segmented the white and gray matter and CSF using FAST4 (Zhang, Brady, & Smith, 2001). The resulting volume counts on these compartments were directly used for the analysis.

For the study on personality we performed voxel-based morphometry carried out with FSL (Smith et al., 2004). For this study the structural images were brain-extracted (Smith, 2002). Next, tissue-type segmentation was carried out using FAST4 (Zhang, Brady, & Smith, 2001). The so-obtained gray matter partial volumes were then aligned to MNI152 standard space using the affine registration. The resulting images were averaged to create a study-specific template, to which the native gray matter images were then non-linearly reregistered with a method that uses a b-spline representation of the registration warp field (Andersson, Jenkinson, & Smith, 2003). The registered partial volume images were modulated (to correct for local expansion or contraction) by dividing by the Jacobian of the warp field. The modulated segmented images were smoothed with an isotropic Gaussian kernel with a sigma of 4 mm. This procedure was applied to the first and second T1 scans separately, creating to independent data sets. The data set was used to identify regions of interest that explained variance in the overall NEO-PI questionnaire (t-test over the demeaned five factors) using voxelwise permutation-based nonparametric testing. From this we obtained 12 regions of interest (ROI) that we used to extract values in these ROIs from the second (independent) dataset. ROIs were extracted if at least 200 connected voxels surpassed a threshold of $p < .01$. 


Alternative Perspectives in Philosophy of Mind and Their Relationship to Structural Equation Models in Psychology

Richard P. Bagozzi
Ross School of Business, University of Michigan, Ann Arbor, Michigan

Kievit and colleagues (this issue) consider the relationship between "neurological and behavioral measures" within the context of two approaches from the philosophy of mind: identity theory and supervenience theory. Kievit et al. assert that "the reduction problem . . . is a measurement problem" and "identity theory and supervenience theory, can be easily translated into psychometric models" (p. 67).

The relationship between different perspectives in the philosophy of mind and psychometric models is a very important one, and to the best of my knowledge, Kievit et al. are the first to give this issue a detailed philosophical, statistical, and psychological analysis. I think their treatment is a remarkable one, and my comments are intended to be taken as broadening and deepening their presentation in what I would think will become an unfolding dialogue in the literature referring back to Kievit et al.'s seminal article.

Philosophy of Mind Foundations

Psychologists tend to consciously or unconsciously follow one of three approaches with regard to philosophy of mind foundations. Neuroscientists are the most explicit followers of physicalism:

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)

There are at least two philosophy of mind perspectives that have been espoused as fitting neuroscientific approaches to psychology: eliminativism and identity theory (see Figure 1). Both are types of reductionism, where

the basic question of reduction is whether the properties, concepts, explanations, or methods from one scientific domain (typically at higher levels of organization) can be deduced from or explained by the properties, concepts, explanations, or methods from another domain of science (typically one about lower level of organization). (Brigandt & Love, 2008, para. 1)

Eliminativism is graphically depicted in the left of Panel A in Figure 1. Eliminativism "is the radical claim that our ordinary, common-sense understanding of the mind is deeply wrong and that some or all of the mental states posited by common-sense do not actually exist" (Ramsey, 2007, para. 1). Eliminativists in psychology assert that there are no mental states or events but rather only neural, hormonal, and related physical brain processes or states. Leading explications of such neuroscientific eliminativism can be found in P. M. Churchland (1981), P. S. Churchland (1986), and Stich (1983). I have chosen to represent each of the philosophy of mind characterizations in Figure 1 through "causal" diagrams rather than focusing on purely measurement relations as done in Kievit et al. for the identity theory and the supervenience theory, for reasons that become clear next. The diagrammatic conventions that I follow are now common practice in the philosophy of mind literature (see, e.g., Kim, 1996, pp. 149–152). Thus for the eliminativism depiction, Figure 1 shows one physical property $P_1$ influencing another physical property $P_2$, where $P_1$ and $P_2$ in the neuroscience case might be two neural states. Under eliminativism, there are no mental states or properties, only physical brain states or properties. An example of $P_1 \rightarrow P_2$ might be a neural state of pain causing the firing of a motor neuron linked ultimately to a muscle contraction in response to the cause of the pain.

The identity theory of mind "holds that states and processes of mind are identical to states and processes of the brain" (Smart, 2007). In other words, mental properties are identical with physical properties. Figure 1 shows that two mental properties, $M_1$ and $M_2$, are linked, respectively, to two physical properties, $P_1$ and $P_2$. For example, the processing of immediate versus delayed rewards, analogous to impulsive versus planned decisions, occurs in different regions of the brain. Perception of immediate rewards ($M_1, P_1$) leads to activation of the nucleus accumbens and the posterior cingulated cortex during choice ($M_2, P_2$), whereas perception of delayed rewards ($M_1, P_1$) leads to activation of the lateral prefrontal and inferior
parietal corticies (M₂, P₂; McClure et al., 2004). Under identity theory, a mental property is a physical property, and causal relations if any are taken to occur between physical properties. Note that eliminativism denies that there are mental properties; by contrast, the identity theory claims that mental properties are physical properties. At any particular time in scientific progress, a mental property, under the identity theory, may not yet be found to have a physical realizer. So identity theory does not, unlike eliminativism, reject mental properties per se. It is just that mental and physical properties are (eventually) believed (to be shown) to be identical.

Some philosophers and presumably many psychologists might find eliminativism and the identity theory disconcerting. That is, many scholars believe that mental properties operate causally to influence other mental properties and even bodily behavior. One philosopher dramatizes the issue as follows:

If it isn’t literally true that my writing is causally responsible for my searching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying . . . , if none of that is literally true, then practically everything I believe about anything is false and it’s the end of the world. (Fodor, 1989, p. 77)

Kim (2000) noted that

the possibility of psychology as a theoretical science capable of generating law-based explanations of human behavior depends on the reality of mental causation: mental phenomena must be capable of functioning as indispensable links in causal claims leading to physical behavior. (p. 31)

As Kim further noted, the very possibility of human agency requires that our mental states have causal
effects in the physical world and our ability to know presupposes the reality of mental causation. See also Mele (1992, 2003, 2009) for an account claiming that psychological explanation rests on mental causation. In contrast to belief in mental causation, epiphenomenalism maintains that “mental events are caused by physical events in the brain, but have no effects upon any physical events” (W. Robinson, 2007, para. 1). A leading exponent of this latter point of view in psychology and one denying mental causation can be found in Wegner (2002).

Belief in mental causation finds its most developed form perhaps in folk psychology (also known as naïve psychology or commonsense psychology and sometimes termed theory of mind). Folk psychology has been used in three distinct senses to refer to cognitive capacities to predict and explain behavior, a theory of mental causation, and an emergent property dualism. A final point to mention with regard to folk psychology, where we have termed the relationship in Figure 1, “mental causation,” is that mental causation should not be confused with the narrower term, “supervenient cause,” proposed by Kim (1996, p. 151). More on this later.

The third philosophy of mind approach frequently taken in psychology is functionalism. Functionalism is the doctrine that what makes something a thought, desire, pain (or any other type of mental state) depends not on its internal constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part. More precisely, functionalist theories take the identity of a mental state to be determined by its causal relations to sensory stimulations, other mental states, and behavior. (Levin, 2009, section 1, para. 1)

Most research in the experimental tradition in psychology follows a functionalist perspective. Indeed, most research in neuroscience and folk psychology conforms to functionalism. Panel C in Figure 1 shows a schematic of a typical psychological experiment under functionalism. This is a 2 (regulatory orientation: promotion vs. prevention focus) × 3 (implementation intentions: control vs. promotion focus vs. prevention focus intentions) × 2 (habit: strong vs. weak) randomized experimental design (see Tam, Bagozzi, & Spangjol, 2010). Notice that behavior at the left of panel C in Figure 1 consists of manipulated (primed) and measured regulatory focus (P1), primed promotion-focus and prevention-focus implementation intentions plus no implementation intention control (P2), and measured snacking habits (P3); the main dependent variable is snacking behavior recorded over a 2-day period (P4) following the manipulations. Four mental properties function between the behavioral manipulations (P1–P3) and the subsequent behavior (P4). These are the proposed mental states of regulatory focus (M1) and implementation situations (M2), which together define regulatory fit (M3), and habit (M4), which is proposed to interact with M3 to influence P4.

In words, the results of the experiment, under planned contrasts, showed that participants with strong unhealthy snacking habits snacked healthier when they formed implementation intentions that fit their regulatory orientation (i.e., promotion-focused
participants with promotion implementation intentions or prevention-focused participants with prevention implementation intentions) than when they formed implementation intentions that did not fit their regulatory orientation (i.e., promotion-focused participants with prevention implementation intentions or prevention-focused participants with promotion implementation intentions) or no implementation intentions at all. For participants with strong unhealthy snacking habits, making implementation intentions that did not fit their regulatory orientation or not making any implementation intentions at all resulted in similar snacking behavior. Planned contrasts also showed that participants with weak unhealthy snacking habits snacked healthier when they formed implementation intentions (either fitting or not fitting their regulatory focus) than the no implementation formed condition. For people with weak unhealthy snacking habits, forming implementation intentions that either fit or did not fit their regulatory orientation resulted in similar snacking behavior. For a graphic presentation of these findings, see Figure 1 in Tam et al. (2010).

I have presented an extended example of functionalism because it is so pervasive and fundamental to psychology. Mental states under functionalism are characterized by the roles they play in producing behavior. Functionalism also can accommodate multiple realization in the sense that it can represent a mental property as a higher level entity connected to one or more realizers at a lower level. Finally, as I show next, functionalism can be combined with supervenience to accomplish a kind of reductionism. Now we are in a position to more closely examine the treatments of identity theory and supervenience theory by Kievit et al.

The Identity Theory

The identity theory maintains that mental properties are the same as (identical with) physical properties. For the case where psychological processes are studied by neuroscientists, the physical properties constitute such instances as the firing of neurons, the operation of hormones, and the activation of brain regions. Kievit et al. (this issue) conjecture that “identity theory is implicitly assumed in most cognitive neuroscientific work” (p. 74), to which I would add that some neuroscientists seem to follow eliminativism as well. But functionalism and supervenience, among other philosophy of mind frameworks, can be shown to fit the neuroscientific study of psychological processes, too, I claim (see next).

Kievit et al. (this issue) assert that “the reflective [structural equation measurement] model can be used to provide an empirical test of the identity hypothesis” (p. 74). The reflective model they proposed and illustrated consists of a single factor with four measures of intelligence (composed of self-reports to items in four subscales of the Wechsler Adult Intelligence Scale) and three physiological brain measures (white matter, gray matter density, and cerebrospinal fluid volume). The data fit the model poorly, and on this basis Kievit et al. conclude that identity theory must be rejected.

The single-factor reflective model is an apt test of the identity theory hypotheses, but I think a somewhat more informative test is provided by the two-factor model shown in Figure 2. Here, taking the illustration performed by Kievit et al. as an example, I suggest that separate factors for mental properties and for physical properties be specified, with the correlation between the factors estimated. This permits the researcher the
opportunity to explicitly estimate the degree of correspondence \((\phi)\) between mental and physical properties. Given the inevitable presence of measurement error in measures of mental and physiological states, it seems more reasonable to expect something less than perfect correspondence between mental and physical observations in any particular empirical study. The single-factor model proposed by Kievit et al. yields a yes–no test of unidimensionality but does not reveal to what degree correspondence is achieved. I applied the model in Figure 2 to the data analyzed by Kievit et al. and found a good fit overall, \(\chi^2(13) = 17.27, p \sim .19\) (root mean square error of approximation = .062, nonnormed fit index = .97, comparative fit index = .98, standardized root mean square residual = .058). Also the correlation between the two factors was .41, which might be considered "moderately high." If the identity hypothesis is to be sustained, one would like to see a high correlation between the factors, ideally a standardized value for \(\phi\) approaching 1.00. A formal test of whether the estimated \(\phi\) value is indistinguishable from 1.00 can be performed by comparing the findings for the model shown in Figure 2, as is, to a model with \(\phi\) fixed to 1.00. For the data at hand, the model with \(\phi\) fixed to 1.00 gives \(\chi^2(14) = 50.95\), so a test of whether \(\phi = 1.00\) is given by

\[
\chi^2_d(1) = 33.68, \ p < .001,
\]

which shows that \(\phi = .41\) is significantly less than 1.00.

There is one sense that the data used to test for the identity theory hypothesis are not up to the task. The correlations amongst the proposed measures of intelligence suggest poor convergent validity in that the six correlations amongst the indicators range from .20 to .47; the estimated factor loadings for the four intelligence measures confirm low levels of convergent validity: .61, .72, .53, and .49, which shows that trait variance in the four items ranges from 24% to 52%. The data for the second empirical analysis done by Kievit et al. also have the limitation of showing poor convergent validity for the measures of conscientiousness. Here the correlations range from about -.18 to .61 and average only .14, making it difficult to claim that the 11 measures indicate a unidimensional scale as required by the reflective model.

Actually there is a sense in which both the single-factor model proposed by Kievit et al. (this issue) and the one I proposed in Figure 2 is both too demanding, and not demanding enough, to test the identity theory hypotheses. Both models assume that there is no systematic, method error, but to the extent that such error is present, it may not be reasonable to expect the models to fit well and for strict identity to be achieved. Alternatively, even if the identity hypothesis were to be sustained under the two models, unknown method bias could account for or obscure excessive convergent validity, thereby leading to a specious basis for claiming strict identity.

What to do? The only solution that I can envision is to explicitly model trait, method, and error variance. A good way to see this is to take the example of the psychological state or event of empathy. Classically, empathy in psychology has been conceived to exhibit three components: perspective taking (putting oneself in the place of others), empathetic concern (feeling compassion for another person), and self-other differentiation (keeping from feeling excessive personal distress; Eisenberg, 2000; Iacoboni, 2009). Imagine that for each of these components we have for a sample of 150 people, say, measures of self-report, measures by an expert on each of the people (e.g., by a clinician), and fMRI measures of relevant brain regions. Relevant brain regions by use of fMRI might be in the medial prefrontal cortex and inferior parietal lobule for perspective taking (Decety & Lamm, 2006), the insula, the anterior medial cingulated cortex, and amygdale for empathetic concern (Decety & Lamm, 2006; Lamm, Batson, & Decety, 2007), and the precuneus for self-other differentiation (Cavanna & Trimble, 2006).

Figure 3 shows a structural equation model that can be used to test the convergent and discriminant validity of the measures of the components. Notice that each component, each factor, has one self-report, one expert-report, and one fMRI indicator measuring it. Convergent validity will be established to the degree that the overall model fits well and the factor loadings are high (i.e., \(\lambda_1-\lambda_3\) are high for \(M_1\), \(\lambda_4-\lambda_6\) are high for \(M_2\), and \(\lambda_7-\lambda_9\) are high for \(M_3\)). Discriminant validity will be attained to the extent that \(\phi_{21}, \phi_{31}, \phi_{32}\) are statistically less than 1.00. Measurement error will be captured by the variance in the error terms \((\varepsilon_{1-\varepsilon_{9}})\), and method bias will occur to the extent that \(\lambda_{10-\lambda_{18}}\) are large and significant. One wants \(\lambda_1-\lambda_9\) to be high and \(\theta_{1-\theta_{9}}\) and \(\lambda_{10-\lambda_{18}}\) to be small in magnitude.

The model in Figure 3 permits a test of strict and practical identity. For each component factor, mental and physical measures are hypothesized. In principle, strict identity would be demonstrated to the degree that \(\lambda_1-\lambda_9\) are high and error variance and method effects are low. Of course, even with some measurement error and method bias, practical identity might be established. The important point is that any actual test of the identity hypothesis should take into account the possibilities of measurement error and method bias. Self-report and expert-report method bias may be relatively simple in origin and nature yet significant in magnitude. Neurological measures may contain relatively complex and multiple sources of method bias and also might be significant in magnitude. For example, common fMRI measures entail multiple
anatomical/structural scans during which subjects are engaged in some experimental task. Data from subjects might also include processing of visual stimuli via mirrors or auditory stimuli and finger presses of buttons. Recordings of Blood Oxygen level Dependent signals are made, and data from a particular voxel constitute a time series of Blood Oxygen level Dependent signals from 500 to 1,500 time points, typically. Each of tens of thousands of voxel time series may be analyzed. Further, prior to analysis, such preprocessing steps as filtering of voxels outside the brain, taking into account sequential effects, correcting for motion, normalization of data to take into account differences in brain size and shape, and spatial smoothing will be done. Analyses involve fitting a general linear model to each voxel time series, statistical parametric map construction, correcting for temporal auto correlation, and use of Bonferroni or other corrections for multiple comparisons during model fitting. With a host of multiple apparatuses used to collect and process data, one can see that the likelihood of method bias is very real.

A final point. I have used the additive trait-method-error model in Figure 3 to make the aforementioned points. But there may be other models more appropriate to model sources of trait, method, and error, depending on the particular study at hand. For example, traits and methods might statistically interact to produce variance in measures, such that differential attenuation or augmentation of effects occur. This is the direct product model, but other models exist as well. For an overview of various approaches to modeling systematic and random error, see Bagozzi (2011).

The Supervenience Theory

Supervenience was proposed as a way to interpret the relationship between properties conceived at a higher level with properties at a lower level. Kim (2005) defined mind–body supervenience as follows:

Mental properties strongly supervene on physical/biological properties. That is, if any system s instantiates a mental property M at t, there necessarily exists a physical property P such that s instantiates P at t, and necessarily anything instantiating P at any time instantiates M at that time. (p. 33)

Kievit et al. (this issue) believe that a supervenience conceptualization of psychological processes “is consistent with a specific implementation of the formative model, called the MIMIC (for Multiple Indicators, Multiple Causes) model” and “the restrictions and characteristics of the strong supervenience thesis and the formative model are identical” (p. 74). I agree that the MIMIC model can be used to operationalize one sense of the supervenience thesis but at the same time argue that using the MIMIC model has some limitations as well as unfortunate implications. I also maintain that reflective models can be used to operationalize the supervenience thesis.

To clarify my position and point out differences with the treatment presented by Kievit et al., I offer the following. It is easy to be confused by claims made in the philosophy literature and to introduce ambiguity into psychometric implementations of supervenience theory because the relevant treatments in the literature are often unclear, underdeveloped, and
contradictory, and in any case are not done with measurement issues in mind. Kievit et al. take the relationship between a latent psychological attribute and neurological indicators as one of determination. Neurological indicators determine psychological attributes. This is understandable, for even Kim, perhaps the leading supervenience theorist, has used similar language. For example, Kim (1996) noted that “the assumption that is widely shared by physicalists is that higher-level properties are in some sense dependent on, or determined by, their lower-level properties” (p. 222). Likewise, Kim (2000) said, “It is customary to associate supervenience with the idea of dependence or determination: if the mental supervenes on the physical, the mental is dependent on the physical, or the physical determines the mental” (p. 11).

But I think the meaning of dependence here is not the same as in the formative relationship that Kievit et al. (this issue) propose, which they interpret as a causal relationship: “a conceptualization of the theoretical attribute (latent variable) as being in some way causally dependent on its indicators” (p. 72). Kim (2005) reserved the term causal relationship as one describing how one physical property or event, \(P_1\), influences a second physical property or event, \(P_2\). The supervenience relationship is between a mental property, \(M\), and its corresponding physical property, \(P\), and is not a causal relationship under physicalism.

How should one interpret the supervenience relationship? Let us return to Kim (2005):

I take supervenience as an ontological thesis involving the idea of dependence—a sense of dependence that justifies saying that a mental property is instantiated in a given organism at a time because, or in virtue of the fact that, one of its physical “base” properties is instantiated by the organism at that time. Supervenience, therefore, is not a mere claim of covariation between mental and physical properties; it includes a claim of existential dependence of the mental on the physical. (p. 34)

Notice that Kim takes the dependence relationship in an existential sense, where a mental property is instantiated “in virtue of the fact that” its corresponding physical base property is instantiated. Elsewhere, Kim (2000) asserted that determination means “that the mental nature of a thing is entirely fixed [italics added] by its physical nature” (p. 11). Kim (2005) also noted that \(P\) and \(M\) in the definition of mind-body supervenience “occur precisely at the same time” (p. 42), but most conceptualizations of causation stipulate that a cause precedes its effect in time (e.g., Audi, 1995, pp. 110–112). I conclude that the supervenience relationship is not a causal relationship, under Kim’s description, but leave open the possibility of concurrent causation in principle (which, however, does not fit most interpretations of causation by psychologists who presume a manipulation theory of causation).

Another contributing factor to ambiguous interpretation of the nature of the relationship between mental properties and physical properties can be seen in the common practice of interpreting the relationships between latent variables and their indicators, under structural equation models, as causal relationships. I believe that this interpretation is ill-founded. From my point of view, the relationships between latent variables and their indicators should be taken as measurement, not causal, relationships. For decades, since the inception of structural equation models, we have uncritically and mistakenly taken the arrows connecting latent variable and their indicators as causal paths. But this seems to have arisen as a heuristic and practical convenience. Heuristically, it is helpful to think of whatever a latent variable stands for as causing variation in indicators, under the reflective model, or as being caused by its indicators, under the formative model. Such a heuristic is a mental device useful in helping one specify measurement relationships. Practically, the arrows linking latent variables and their indicators serve to represent functional relationships in a measurement sense and to enable one to write equations implied by the measurement relationships. I maintain that this measurement interpretation of reflective models allows one to use it to implement supervenience. See the next major section of this article next.

Of all the possible formative models (see Bagozzi, 2011), the MIMIC model proposed by Kievit et al. provides the least ambiguous representation of a formative relationship. But it should be recognized that the latent formative variable in the MIMIC model is difficult to interpret as a mental or physical property. It is neither solely a latent physical property factor formed by the physical indicators nor solely a latent mental property factor reflected by the psychological measures. Rather it is best to think of the latent formative factor as a mathematical operator, performing a transformation of a linear combination of the formative indicators into a linear combination of the reflective indicators. This is analogous to the operation performed in canonical correlation analysis. The MIMIC model does constitute a specific hypotheses relating physical to psychological measures, but the correspondence of this model to supervenience theory is not as exact as claimed by Kievit et al. That is, the MIMIC model relates measures of physical properties to measures of mental properties, but supervenience is at best implied in the model in that supervenience is defined to be between mental properties and physical properties, and the latent formative factor is neither a precise representation of either the mental or physical property. Rather, the factor is indeterminant in meaning because of its joint dependence on the measures of both the physical and mental.
properties. Other drawbacks with the MIMIC model and other formative models are the inability to make claims about internal consistency reliability and construct validity and the difficulty in talking about generalizability and making comparisons amongst these models in empirical research.

The final issue I wish to consider with regard to supervenience is at the heart of the psychometric approach to the reduction problem and the presentation in Kievit et al. Kim (2005) maintained that, under physicalism, one must choose between supervenience and reduction. Figure 4 displays the choice. Panel A shows how supervenience can be depicted: two mental properties, M and M*, are linked to their respective physical realizers, P and P*, and the causal relationship seemingly occurring between M and M* actually happens at the physical level between P and P*. Panel B shows the reduction argument diagrammatically, where the putative causal path from M to M* is really one and the same with the causal path from P to P*. The key point under reduction is that M = P and M* = P*, and we have only one causal relation. Under supervenience, M \neq P, M* \neq P*.

How might we think of the possibility of reduction under supervenience theory? Kim (2005) argued that, with the possible exception of qualia, mental properties are functionalizable and can “be defined or characterized in terms of their causal work” (p. 165). Kim (2005) used functionalist arguments to explain how one can accomplish a physical reduction of something mental and proposes three steps:

1. The first is a conceptual step of interpreting, or reinterpreting, the property to be reduced as a functional property, that is, in terms of the causal work it is supposed to perform. Once this has been done, scientific work can begin in search of the “realizers” of the functional property—that is, the mechanisms or properties that actually perform the specified causal work.

2. The third step consists in developing an explanation at the lower, reductive level of how these mechanisms perform the assigned causal work. (p. 164)

This can be accomplished under supervenience theory, and reflective structural equation models can be used to conduct empirical tests of hypotheses derived from theories specified under supervenience theory.

A Holistic Representation of Philosophy of Mind Frameworks Under Structural Equation Models

To better understand the relationship between measures of mental properties and measures of physical properties, consider Figure 5, Panel A. Notice that mental properties (M) and physical properties (P) are linked through mind–body relations, of which there are a number of possibilities in philosophy that fit various points of view in psychology. Perhaps the oldest perspective is given under substance dualism (e.g., Cartesianism) where substances in the world are divided into mental (immortal) and nonmental (material) varieties, and both interact causally in mind–body relations. Nowadays, philosophers no longer follow Cartesianism (or other variants of substance


Figure 5. A holistic representation of philosophy of mind frameworks and their relationship to structural equation models.
dualism), but it is not difficult finding some psychologists who subscribe to substance dualism, at least implicitly, in that they accept that the mind or selves interact with the body. This point of view is difficult to defend philosophically, and I include it herein merely for completeness and to point out that some psychologists have not abandoned the perspective. More common in psychology is to be found kinds of property dualism wherein the claim is made that mental properties are irreducible to physical ones, although they are in some sense realizable physically (Robb, 2008; H. Robinson, 2007). A property dualist orientation would seem to fit folk psychology, mental causation, emergentism, and perhaps downward causation (i.e., the view that mental properties cause physical ones). Functionalist accounts might be adapted to property dualism, too. The other relatively common mind–body relation presumed in psychology might be subsumed under the label physicalism. Here eliminativism, identity theory, and supervenience are alternatives under physicalist points of view. Causation is asserted to occur between physical properties, and under supervenience mental properties are linked in the way quoted previously by Kim (2005).

It can be seen in Figure 5, Panel A, that measures of mental properties are connected to measures of physical properties through the mind–body relation. It is important to realize that such measures and the theoretical and conceptual properties that they are linked to are distinct variables. Psychologists have not always acknowledged this distinction and therefore have ignored the implications of random and systematic (e.g., method) errors when implementing structural equation models. The mind–body relation has consequences for representing the relationship between measures of mental properties and measures of physical properties that should be taken into account explicitly in any structural equation model specification. This requires that proper measurement relations (reflective or formative) be specified for measures of both mental and physical properties. The different mind–body relations imply different specifications linking the latent variables corresponding to mental and physical properties. Under eliminativism only latent physical property factors are specified. Under the identity theory, both latent mental and physical property factors are specified but are expected to be highly correlated after measurement error and method biases are taken into account. Under folk psychology, most factors will correspond to mental properties, although physical property factors could be specified as well, but here mental and physical property factors are linked via mental causation. Under functionalism, mental property factors relate to each other through mental causation but achieve their meaning through the intermediary roles they play in connecting behavioral or physical phenomena to other behavioral or physical phenomena. Under property dualism, mental properties cannot be reduced to physical properties, and unlike under supervenience, the relationship between mental properties and physical properties seems to be one of covariation with, depending on the particular property dualism espoused, something added to the covariation to permit a proper grounding of the mental on the physical. Under supervenience, the mind–body relation has the existential, noncausal meaning proposed by Kim (2000, 2005). Also, supervenience denies that there is mental causation and instead claims that causation occurs only between physical properties.

How might one represent the implications of the general philosophy of mind representation in Figure 5, Panel A, for structural equation models? Figure 5, Panel B, presents an elliptical possibility that subsumes many of the philosophy of mind perspectives previously mentioned. For purposes of discussion, I assume that the researcher wishes to test the hypothesis that T1, influences theoretical property, T2, with γ3 the inferred (i.e., estimated) parameter summarizing the result of testing this hypothesis statistically. Under eliminativism, there are no mental properties (M1), so the physical properties (P1) collapse into their respective theoretical properties, and γ5 is estimated between T1/P1 and T2/P2, where the measures used to accomplish this are p1–p4. Under identity theory, T1 and T2 are higher level theoretical properties attached to their respective lower level properties, M1 and P1, and M2 and P2. To the extent that the identity theory is sustained, γ1 and γ2 will be high in value, and β1 and β2 will be high in value, and ζ1, ζ2, ζ4 and ζ5 will be low in value. Under folk psychology, either M1 or P1 and either M2 or P2 will collapse into T1 and T2, respectively, such that γ3 will capture mental causation between M1 → M2, P1 → M2, or M1 → P2 (P1 does not influence P2 under folk psychology but does under eliminativism). Under both property dualism and supervenience, M1 and P1 load on T1, and M2 and P2 load on T2, so to speak, but the relationship between M1 and P1 and the relationship between M2 and P2 have different meanings. The conceptual meanings of the mind–body relations are implicit for property dualism and supervenience in Figure 5, Panel B. That is, under both conceptual meanings of the mind–body relation for property dualism and supervenience, we model the empirical implications of the conceptual meaning, which are similar in both cases. For example, under both perspectives, the measures m1, m2, p1, and p2 will share considerable common variance empirically because they relate to M1 and P1, respectively, through measurement relations, and the variances are produced by T1, T1 hypothesizes that M1 and P1 are highly correlated, as a consequence of their respective measures, and the fact that M1 and P1 are linked formally via the mind–body relation under property dualism and supervenience. Although the mind–body relations differ here, they have similar
Comment on the Relationship of Personality Theory and Neuroscience

The two main empirical analyses in Kievit et al. (this issue) were performed on individual difference variables (i.e., intelligence and conscientiousness). In addition to the identity theory and supervenience approaches that they took by use of reflective and formative models, it is possible to study personality variables and the relationship between neurological data and behavioral/psychological data by use of structural equation models and neuroscience experiments, under a functionalist perspective. Here I give an example.

Dietvorst et al. (2009) studied the relationship between a new theory of mind (ToM) scale and neurological measures corresponding to brain regions related to ToM processes. Based on research on autism, communication, and related areas, a scale development procedure was conducted that generated 33 potential items. Further psychometric procedures were applied, and a four-dimensional scale with 13 items was eventually derived. The four ToM dimensions covered important interpersonal mentalizing skills: building rapport with an interaction partner, detecting nonverbal cues, taking a bird’s-eye view of one’s interactions, and shaping the interaction. Three separate studies were then conducted on three samples of respondents to investigate construct validity. Structural equation models were employed across the studies to demonstrate convergent validity of items, discriminant validity of items across the four subscales, discriminant validity of items with items measuring other scales, criterion-related validity of items, predictive validity, and nomological validity. One of the three studies employed a multitrait-multimethod design.

After confirming construct validity of the items of the new ToM scale, a fourth neuroscience study was conducted. The ToM scale was administered to a sample of 132 sales managers, and the 10 highest and 10 lowest scorers on the scale were selected to participate in a fMRI investigation. Participants engaged in three experimental conditions: interpersonal mentalizing, process, and unlinked sentences tasks. The participants listened to five stories under each of the three conditions, which were presented in one of two counterbalances orders. The interpersonal mentalizing task was the critical condition, in which the cognitive task involved the use of ToM to understand why and how protagonists in the stories interact. The process condition served as a closely matching control condition, in which the cognitive task involved nearly the same cognitive processes as the interpersonal mentalizing condition, except that the stories did not require ToM to comprehend how and why the protagonists interact. In the unlinked sentences tasks, participants listened to a series of sentences not making for a coherent story. This is a baseline control condition in which the cognitive task involved simple language and memory processing. Each experimental condition was followed by a question asked of participants to answer silently. Each of the 15 stories lasted 33 to 36 s, and each participant was given 6 s to reflect on an answer to each question posed.

Based on research in the autism and interpersonal mentalizing literatures, it was predicted that the medial prefrontal cortex (MPFC), left and right temporoparietal junction (TPJ), and left and right temporal poles (TP) would be activated at higher levels for the interpersonal mentalizing versus process and unlinked sentences conditions. The findings showed that more activation was indeed achieved of the MPFC and TPJ but not the TP for the interpersonal mentalizing versus control conditions. With regard to the TP, this brain region was found to be activated highly for all conditions. A post hoc interpretation was given for this latter finding based on the role of script-based thinking, which is believed to occur in the TP region and which is characteristic of the training and performance requirements of sales managers. Finally, correlational analyses revealed that the ToM scale correlated .69 with the right TPJ, .69 with the MPFC, and .61 with the left TPJ signal intensity scores.

The Dietvorst et al. (2009) study represents an alternative way to investigate the relationship between behavioral/psychological data and neurological data. It does this within the framework of functionalism. The individual difference ToM scale is used to make predictions of cognitive processes under experimental conditions, where brain areas known to relate to interpersonal mentalizing are also predicted to reflect the processes under the ToM and interpersonal mentalizing task. This investigation shows a holistic functionalist approach to the study of the relations of behavioral/psychological and neurological data.
Conclusion

Philosophy of mind frameworks in general, and the identity theory and supervenience theory in particular, are not easily translated into psychometric models. Philosophy of mind frameworks deal with the logical or conceptual foundations of concepts (e.g., mental properties and physical properties). Structural equation models are statistical methods for testing empirical predictions. The bridge between conceptual foundations of concepts and statistical specification of structural equation models requires that one attempt also to ascertain the implications of mind–body relations for latent variables and their relationships to measures and other latent variables. This means, for example, specifying the nature of the measurement relationship between latent variables and their measures, in the light of mind–body relations. The meaning of the relationship between a latent variable and its measures does not inhere solely in the factor loading parameters inferred from data. Rather, the relationship includes meaning going beyond the empirically derived parameter to encompass logical or conceptual meaning, such as resides, for instance, in criteria used to deduce a measure from a theoretical latent variable and the conceptual meaning linking a latent variable to its measures (Bagozzi, 2011). Such meaning goes beyond the empirical content of a factor loading and is in a sense surplus meaning or auxiliary meaning. Further, as suggested in Figure 5, Panel A, the relationship between measures of mental properties and measures of physical properties includes two measurement relations plus a mind–body relation. These criteria go beyond the meaning of empirical parameters in structural equation models and introduce conceptual meaning into the interpretation of meaning of factor loadings. Substantive researchers do not always develop the conceptual meaning underlying measurement relations, but they are implied by or implicit in the theoretical development of hypotheses and deserve explicit consideration. The relationship between measures of mental properties and measures of physical properties is made complex and difficult to precisely specify by the different mind–body interpretations and their implications for measurement relations. We should not ignore the conceptual meanings and their implications for measurement. Kievit et al. and I have considered some of the issues previously mentioned, but more remains to be done.

Likewise, the reduction problem is not solely a measurement problem. To be sure, measurement issues are important concerns for reduction. But as suggested in Kim’s (2005) functionalist approach to reduction previously discussed, reduction also involves conceptual interpretation of a mental property to be reduced in terms of the role it plays in a functionalist model and discovery of physical realizers of the functional property and development of an explanation at the reductive level. These operations go beyond measurement issues or criteria. Approaches other than functionalism apply to the reduction issue, too, of course.

Kievit et al. (this issue) provide an original and important impetus to study the reduction problem and to the simultaneous analysis of behavioral/psychological and neurological data. Their treatment of the identity theory and supervenience theory and its implementation through reflective and formative models are exemplary. I believe their article is the first to bring together the intersection of psychology, psychometrics, and philosophy as they apply to the relation between behavioral/psychological and neurological data.

Note

Address correspondence to Richard P. Bagozzi, Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109-1234. E-mail: bagozzi@umich.edu

References


response inhibition based on the prospective relationship between craving and subsequent smoking. For example, a subject who reported high levels of craving at 10 a.m. and then reported having smoked two cigarettes at the following signal (at 12 p.m.) was treated as having had a response inhibition failure between 10 a.m. and 12 p.m. These self-reports of smoking behavior were corroborated with two biological measures of cigarette smoking (urinary cotinine and exhaled carbon monoxide).

Using hierarchical linear modeling, we estimated the time-lagged within-day slope between craving (at time $t$) and smoking (at time $t+1$) for each participant. Estimates of response inhibition-related neural activation from the baseline session were entered into the model and allowed to moderate the craving-smoking slope. Results showed that our neural measure of response inhibition from the baseline session significantly moderated the behavioral indicator of response inhibition that was derived using experience-sampling data (Figure 2). Task performance during the baseline scan (a presumptive P-indicator) was relatively high and uniform across subjects and consequently did not relate to the craving-smoking link (a relatively distal behavioral indicator).

These results are interesting not only because they demonstrate the ecological validity of neural measures of response inhibition but also because they illustrate a couple of the key points made in this commentary. First, this study shows that real-world behavior can realistically be integrated into traditional cognitive neuroscience studies with meaningful results. Taken together with the statistical models described in the target article, these results point the way forward in bringing behavioral measures back together with P- and N-indicators. Second, the lack of association between the P-indicator (task performance on the go/no-go task at baseline) and the B-indicator (daily craving-smoking link) is consistent with our analysis on distal behavioral measures. Countless other factors might influence behavior when it is observed in situ outside of the laboratory—factors that are important to identify and understand. For example, in another paper on this data set we found that negative mood was an important moderator of the daily craving-smoking link (Berkman, Dickenson, Falk, & Lieberman, 2011). This finding is consistent with a multiple-indicators multiple-causes model of P-indicators to B-indicators and provides further evidence of the value of real-world behavioral data in addition to psychological and neural indicators.

### Health Behavior Change and Persuasion

The discrepancy between stated attitudes about health behavior and actual changes in health behavior continue to be a hot topic of study in health psychology (e.g., Webb & Sheeran, 2006). We sought to address this problem in a series of studies led by Falk by attempting to explain additional variance in health behavior, above and beyond self-report measures, using neuroimaging (Falk, Berkman, Mann, Harrison, & Lieberman, 2010). To use the terminology of Kievit et al., we added N-indicators to an area that traditionally only uses B- and P-indicators.

In a first study, participants were scanned using fMRI while they were shown persuasive messages about increasing sunscreen use (Falk et al., 2010). Sunscreen use was chosen because it is a common health behavior, and our participants were preselected to have weak preexisting attitudes about it. We measured attitudes and intentions about sunscreen use before and immediately after the message exposure (P-indicators), and then measured change in sunscreen use 1 week later (a distal B-indicator). Neural activation in a predefined region of interest in the medial prefrontal cortex (MPFC) uniquely explained about 25% of the variance in behavior change above and beyond self-report. In other words, just as in the study on response inhibition and smoking, part of the neural activation that related to behavior was not explained by

\[ \text{Average } \# \text{ of cigarettes smoked at time point } t+1 \]

\[ \text{Craving at time point } t \text{ within a day} \]

**Figure 2.** The moderating effect of response inhibition at baseline on the association between craving and smoking across 3 weeks of in situ smoking cessation (Berkman et al., in press). Note. Activation in right inferior frontal gyrus (rIFG) in the response inhibition contrast (no-go > go) moderates the relationship between cravings at one time point and smoking at the subsequent time point (log-slope = –0.29, SE = 0.12), $t(2391) = 2.38, p < .05$. Individuals with low activation in rIFG (–1 SD of the mean) in the [no-go > go] contrast showed a strong positive relationship between cravings and subsequent smoking (simple slope (log units) = 0.53), $t(25) = 2.79, p < .01$. Individuals at the mean showed a modest positive relationship (simple slope (log units) = 0.24), $t(25) = 1.20$, ns, and individuals with high activation (+1 SD of the mean) showed no relationship between craving and smoking (simple slope (log units) = –0.04), $t(25) = 0.21$, ns. This analysis controls for the linear decline in smoking across days, the negative quadratic pattern of smoking within each day, and baseline nicotine dependence.
traditional self-report measures of psychological processes. A second study replicated these results using cigarette smoking cessation, a more meaningful and health-relevant behavior (Falk et al., 2011). This study found that neural activation in MPFC during exposure to persuasive messages (quit smoking television advertisements) was predictive of quitting behavior above and beyond self-reported intentions and attitudes.

An important implication of these studies is that the psychological processes that drive behavior may be only partially known or accessible to self-report. As a consequence, any study based purely on P-indicators may be missing important moderating factors. Neural indicators together with behavior can help triangulate missing psychological processes that might be relevant. For example, the fact that activation in MPFC—a region often associated with self-related processing—predicts behavior change above and beyond reported persuasion hints that self-related processing (e.g., a match between the message and current personal goals) might be an important psychological process in health behavior change that people do not typically include in their subjective reports of persuasiveness. Distal measures of behavior may be particularly useful because they can capture subtle situational and personal factors that may be overlooked or inaccessible to introspection or narrow experimental measures of psychological processes. In this way, measures of behavior that occur in the context of ongoing daily experience are critical to building a complete model of the link between neural measures and psychological processes.

Concluding Remarks

Kurt Lewin’s (1943) famous formula states that behavior is a joint function of person and environment. Kievit et al.’s target article exemplifies a trend in recent years in psychology and neuroscience to focus on the “person” piece of the equation in exquisite detail by examining neural and internal mental processes. However, much of this detail has come at the expense of neglecting the other two components—behavior and environment. We have argued here that Lewin is essentially correct in two ways, and modern empirical psychology would benefit from revisiting his theory. First, Lewin (and others) believed that the fundamental goal of psychology should be to explain human behavior and that mental processes are important insofar as they are mediators of that behavior. Second, behavior must be understood as an interaction between the person and the environment. Because the psychological processes that are reflective of the environment can be difficult to assess with self-report (e.g., contextual primes or cue-induced mood), measuring behavior and psychological processes in situ using methods such as experience sampling are critical to obtaining a complete understanding of behavior.

Careful theoretical work is necessary to integrate behavior into existing models of the relationship between neural and psychological processes. Kievit et al. have done much of this work by providing a psychometric framework for empirically comparing models of this relationship. We have extended their model to include behavior and have suggested a few other issues to consider, particularly the distinction between proximal and distal measures of behavior. Of course, there is still considerable work to be done both statistically and theoretically. To give two examples, it would be useful to have an SEM model (similar to Figure 1) that allowed for hierarchical nesting of multiple behavioral measures, or for the latent behavioral construct (in the formative model from psychological to behavioral processes) to be better specified. We hope that the framework described here provides a foundation for future studies to address these and other outstanding issues.

Note

Address correspondence to Elliot T. Berkman, Department of Psychology, University of Oregon, 1227 University of Oregon, Eugene, OR 97403-1227. E-mail: berkman@uoregon.edu

References


HIT on the Psychometric Approach

Daniel Burnston, Benjamin Sheredos, and William Bechtel

Introduction

Traditionally, identity and supervenience have been proposed in philosophy of mind as metaphysical accounts of how mental activities (fully understood, as they might be at the end of science) relate to brain processes. Kievet et al. (this issue) suggest that to be relevant to cognitive neuroscience, these philosophical positions must make empirically testable claims and be evaluated accordingly—they cannot sit on the sidelines, awaiting the hypothetical completion of cognitive neuroscience. We agree with the authors on the importance of rendering these positions relevant to ongoing science. We disagree, however, with their proposal that a metaphysical relationship (identity or supervenience) should “serve as a means to conceptualize a version of the identity theory—heuristic identifications of phenomena such as memory or attention (albeit typically more finely delineated). Any importation of metaphysical theories should be tightly bound up with such explanation of phenomena, not solely with the correlation of variables that results from measurements. In contrast, Kievit et al.’s project seeks to account for correlations between types of measurements via purely formal/mathematical descriptions of their relationship. They then propose to test whether identity or supervenience is better supported by specific models of the correlations between measured variables. This is not in tune with the explanatory endeavors of cognitive neuroscience.

We consider first their treatment of identity theory. The authors propose to formalize identity theory with “reflective models,” in which the states of multiple “indicators” are represented in a structural equation as “caused by” the state of a common latent variable (see Figure 3 and p. 73 of their article). Each indicator is provided a specific value via some measurement. In the authors’ case studies, the values of psychological or “P-indicators” and neurological or “N-indicators” are specified by measurements such as scored performance on a task. The neurological or “N-indicator(s)” are measurements such as brain mass volume or gray matter density. An identity claim is viewed as justified when a reflective model has a high degree of fit to both sets of data; with a high degree of fit, one may infer that one is “measuring the same thing” with both P- and N-indicators (p. 73). It is the structural equation that shows that the causes of the two variables are identical: the cause of each is the value of the latent variable. When the reflective model has a good fit to the data, it is hypothesized that “the indicators measure the same thing,” where this implies (perhaps incorrectly) that “the latent variable or attributes are measured, and phenomena, which are repeatable processes in the world that are to be explained. Measurements can provide an epistemic inroad to phenomena but should not be confused for phenomena themselves. Reading ability is a cognitive phenomenon; response time on the Stroop task is often used as a measure of that phenomenon. Attention is a cognitive phenomenon; success at discriminating targets from distracters is one measure of it. Memory is a cognitive phenomenon; accurate recall is often used to measure it. If one examines any research area in cognitive neuroscience, it is clear that the researchers are trying to explain phenomena such as memory or attention (albeit typically more finely delineated). Any importation of metaphysical theories should be tightly bound up with such explanation of phenomena, not solely with the correlation of variables that results from measurements. In contrast, Kievit et al.’s project seeks to account for correlations between types of measurements via purely formal/mathematical descriptions of their relationship. They then propose to test whether identity or supervenience is better supported by specific models of the correlations between measured variables. This is not in tune with the explanatory endeavors of cognitive neuroscience.

Explanation of Correlations Versus Explanation of Phenomena

To situate our critique, we draw upon Bogen and Woodward’s (1988) distinction between data, which
exists independently of the model specification” (p. 72).

What Kiev et al. offer is a metaphysical account of how two sets of measurements can be correlated. But, as previously noted, cognitive neuroscience is not in the business of accounting for correlated data sets; rather, its goal is to explain cognitive phenomena themselves. The latent variable identified in a formative model does not play a role in this type of explanation. In part, this is because it has no positive description. The authors’ only characterization of the latent variable is as “cause of the indicators.” Yet even this is merely an interpretation: All that is inherent in the model is a formal description of the latent variable’s relation to the indicators (Borsboom, Mellenbergh, & van Heerden, 2003).

The important point to note here is that the latent variable is defined only with respect to this relation, and therefore there is no characterization of any cognitive or brain process at work separate from the indicators. Without such a description, the target of the explanation cannot in fact be the phenomenon of interest but only the data itself. Given that there is a real distinction between measurements and phenomena, the authors’ models, and the latent variables that are supposed to pull metaphysical weight in them, are only accounting for the measurement side of the distinction.1 If we are right in characterizing cognitive neuroscience, and the metaphysical commitments it makes, in terms of explaining phenomena, then the authors’ models are not being invoked to construct a genuine metaphysics of cognitive neuroscience at all.2 Even given a successful reflective model, if the explanatory questions that cognitive neuroscientists actually ask are broached—that is, what is the phenomenon at issue? and how does it come about?—the reflective model has no answer, other than to say that the “indicators [of the phenomenon] can be said to be on equal empirical footing in that they are both assumed to be imperfect reflections of the true state” of the underlying cause (p. 73). We are left waiting for a metaphysical account of cognitive phenomena.

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

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

---

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

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

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

4 We have two concerns that are specific to the treatment of latent variables in formative models. The first involves the metaphysical status of the latent variables. There is an ongoing debate regarding the ontological status of latent variables, in which one of the authors of the target article is a participant (Borsboom, 2008; Borsboom et al., 2003). The status of latent variables in formative models is particularly vexed, because (as admitted in the target article) in formative models, “the latent attribute is defined by the choice of predictors” and “a change of predictors implies a change in the nature of the attribute” (p. 72). As a result, scholars disagree over whether to take a broadly realist or antirealist interpretation of latent variables in formative models. This is a topic one would hope to see further addressed by the authors, because it has serious consequences for a metaphysical interpretation of the model. If it is best to adopt an antirealist approach to the latent variable in formative models, then clearly the metaphysical goals of the authors will be unsatisfied. Second, it is unclear whether the formative model actually captures multiple realizability. The authors state that “two people can have different indicator values but the same position at the latent attribute level. Therefore the position on the theoretical attribute is multiply realizable” (p. 74). This statement is doubly perplexing. First, if the latent variable just is some function of the neural values, it would seem to be itself neural in nature, but traditionally, supervenience theory has concerned the multiple neural realizability of psychological features. The authors speak of the latent variable as if it were psychological, but it is unclear why they feel licensed in doing so. Further, if we assume (in accordance with the model) that the “different indicator values” regard N-indicators, then the claim is true
the formative model is only construed as determining the relationship between measurements, it only accounts for data, and does not explain phenomena. The deficiency of this view, in terms of explanatory import, becomes clear in the authors’ case studies. In one study, they obtain data correlating “intelligence” (defined as measurements on a series of tests) and brain volume (defined as measurements taken using voxel-based morphography, or VBM). They then apply both formative and reflective models to the data and observe that the formative model generates a considerably better fit in correlating these two data sets than the reflective model. The authors thus conclude that intelligence supervenes on brain matter and that the metaphysical relation between them is “solved” via the measurement. This conclusion suggests that the model completes the investigation by providing a satisfactory view of the metaphysics of intelligence. We contend that this is entirely insufficient as a neuroscientific explanation of intelligence, one that no cognitive neuroscientist should (or probably would) be satisfied with. The reason is that we have only accounted for data. Both the nature of intelligence and of the mechanisms underlying it remain a complete mystery. In the next section, we offer a very different view of how to import metaphysics usefully into cognitive neuroscience, which takes a metaphysical posit as a beginning to and a guide toward mechanistic explanation.

Identity Claims and Mechanistic Explanations in Cognitive Neuroscience

As we argued in the previous section, cognitive neuroscience is typically engaged in explaining phenomena. The form of such explanation is typically mechanistic. Recently philosophers of biology have attempted to articulate what mechanisms are and how they figure in explanations. Fundamentally, a mechanism consists of parts performing operations that are organized and coordinated so as to produce the phenomenon of interest; a mechanistic explanation characterizes the responsible mechanism and shows how it could produce the phenomenon (Bechtel & Abrahamsen, 2005; Bechtel & Richardson, 1993/2010; Glennan, 1996; Machamer, Darden, & Craver, 2000). The only if two people’s brains are measured with the very same styles of measurement. Add or subtract to the data and the latent variable changes. Assigning the same numerical value to the latent variables would then not be assigning the “same” value to the “same” latent variable: it would not be multiple realizability.

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

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

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

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

The important thing to notice is how the field has developed in the few years since this controversy emerged. One strategy has been to tease apart the specific contributions of the FFA in relation to the more general object-recognizing features of the ventral visual pathway, of which it is a part (Grill-Spector, Golarai, & Gabrieli, 2008). It is important to note that attempts have also been made to uncover functional subregions within the FFA. Both Grill-Spector, Sayres, and Ress (2006) and Haxby (2006) used high-resolution fMRI to locate particular, interspersed areas within the FFA that are preferentially activated to faces as well as ones that responded preferentially to other categories. Grill-Spector et al. claimed that the FFA is in fact “reliably heterogeneous” as to its structure and function (p. 1177), whereas Haxby suggested that, at a fine-grained level, function in the FFA is “distributed” across several independent subunits. Both views suggest that the original FFA data were the result of averaging over activation from several distinct subunits within the FFA.

From the standpoint of mechanistic research, this development represents an early attempt at decomposing the FFA. After a rough localization of facial recognition to the FFA (through the first identity claim), research is now progressing (through revisions to the first identity claim) to discover the fine-grained mechanisms and subprocesses involved in producing the phenomenon.

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

---

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

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

10For a differing opinion based on a technical objection to the use of high-resolution fMRI, see Baker, Hutchinson, and Kanwisher (2007).
(2007) pursued such a strategy, applying repetitive inhibitory trans-cranial magnetic stimulation (TMS) to the right occipital gyrus, an area prior to the FFA in the ventral stream that had itself previously been implicated in face processing. When TMS was applied, facial discrimination was severely limited, thereby suggesting a multistage model of facial processing, in which activation in the occipital gyrus creates an “initial representation” (Pitcher et al., 2007, p. 1569) of the face. This result is relevant to both strategies just discussed: the further localization of specified functioning (i.e., initial face representation vs. face recognition simpliciter) to different brain areas represents a more fine-grained account of the phenomenon, and a new spur to further decomposition.

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

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

**Levels, Mechanisms, and Identity Claims**

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

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

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

Conclusions

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

Note

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

References


Bridging Token Identity Theory and Supervenience Theory Through Psychological Construction

Lisa Feldman Barrett

Department of Psychology, Northeastern University; and Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts

A psychologist’s task is to discover facts about the mind by measuring responses at the level of a person (e.g., reaction times, perceptions, eye or muscle movements, or bodily changes). A neuroscientist’s task is to make similar discoveries by measuring responses from neurons in a brain (e.g., electrical, magnetic, blood flow or chemical measures related neurons firing). Both psychologists and neuroscientists use ideas (in the form of concepts, categories, and constructs) to transform their measurements into something meaningful. The relation between any set of numbers (reflecting a property of the person, or the activation in a set of neurons, a circuit, or a network) and a psychological construct is a psychometric issue that is formalized as a “measurement model.” The relation is also a philosophical act. Scientists (both neuroscientists and psychologists) who make such inferences, but don’t explicitly declare their measurement models, are still doing philosophy, but they are doing it in stealth, enacting certain assumptions that are left unsaid.

In “Mind the Gap,” Kievit and colleagues (this issue) take the admirable step of trying to unmask the measurement models that lurk within two well-defined traditions for linking the actions of neurons to the actions of people. They translate identity theory and supervenience theory into popular measurement models that exist in psychometric theory using the logic and language of structural equation modeling. By showing that both philosophical approaches can be represented as models that relate measurements to psychological constructs, Kievit et al. lay bare the fact that all measurement questions are also philosophical questions about how variation in numbers hint at or point to reality. They make the powerful point that translating philosophical assumptions into psychometric terms allows both identity theory and supervenience theory to be treated like hypotheses that can be empirically evaluated and compared in more or less a concrete way. The empirical example offered by Kievit et al. (linking intelligence to brain volume) is somewhat simplistic on both the neuroscience and psychological ends of the equation, and the nitty-gritty details of applying an explicit measurement approach to more complex data remains open, but this article represents a big step forward in negotiating the chasm between measures taken at the level of the brain and those taken at the level of the person.

The overall approach is applauded, but a closer look at the details of how Kievit et al. operationalized identity and supervenience theory is in order. In science, as in philosophy, the devil is in the details. In the pages that follow, I highlight a few lurking demons that haunt the Kievit et al. approach. I don’t point out every idea that I take issue with in the article, just as I don’t congratulate every point of agreement. Instead, I focus in on a few key issues in formalizing identity and supervenience theory, with an eye to asking whether they are really all that different in measurement terms, as well as whether standard psychometric models can be used to operationalize each of them equally well. Like Kievit et al., I conclude that a supervenience theory might win the day, but I try to get more specific about a version of supervenience that would successfully bridges the gap between the brain and the mind.

What Is the Correct Measurement Model for Identity Theory?

The first issue to consider is whether Kievit et al. provided the correct measurement formalization for identity theory. Identity theories define the mental in terms of the physical (i.e., they ontologically reduce mental states to states of the nervous system). Right off the bat, this contradicts Kievit et al. claim that neural and psychological measures are on equal footing. As a consequence, I operationalize identity theory slightly differently than do Kievit et al. Also, there are two versions of mind:brain identity: type identity and token identity. When discussing identity theory, Kievit et al. explicated the type version, but the token version is important to consider because it blurs the distinction between identity and supervenience.

The type version of identity theory assumes that psychological kinds are physical kinds (e.g., Armstrong, 1968; Place, 1956; Smart, 1959). It assumes that the mind is populated by abstract categories for different mental faculties (like emotion, memory, perception, intelligence, etc.) and correspondence between mind and brain resides at the level of the abstract...
Type identity theories of emotion (e.g., "basic emotion" models and some appraisal models), for example, assume that certain emotion categories (e.g., fear) can be reduced to the activation of one and only one brain region (e.g., Calder, 2003; Ekman, 1999), one specific neural circuit (e.g., Izard, in press; Panksepp, 1998), or one physiological state (e.g., Ekman & Cordano, in press; Levenson, in press). Sometimes types are conceived of as natural kinds (e.g., Panksepp, 2000) or sets (e.g., as in families of emotion; Ekman, 1992). In the extreme, type identity theories do away with mental concepts altogether because they can be merely redefined in terms of their physical causes (e.g., Feyerabend, 1963).¹

Kievit et al. formalize type identity theory using a measurement model called an effect indicator model (also called a reflective model; Bollen & Lennox, 1991) as Figure 1a. In this kind of a model, the circle represents a hypothetical but not directly observable (i.e., latent) construct that is estimated by observed, measurable variables represented by the squares (Kievit et al. label the squares as “N” and “P” for neural and psychological measures, respectively). In an effect indicator model, the measured variables correlate with each other perfectly (barring measurement error) because they have a common cause (the hypothetical construct). Statistically, their observed correlation is taken as evidence that the hypothetical construct exists (because it cannot be measured directly by its nature or given the limits of existing measurement tools); mathematically, this shared correlation estimates the value of the latent hypothetical construct. The observed measures are said to be indicators or reflections of that hypothetical construct. In Kievit et al.’s view, this hypothetical construction is akin to a psychological faculty, like intelligence (their example), or presumably emotion, memory, perception, and the like.

It is unclear, however, whether an effect indicator model is the correct measurement model for type identity theory, largely because a hypothetical construct might not be needed. If mental states are nothing more than physical states (in this case, states of the nervous system), so that one (the mental) can be ontologically reduced to the other (the physical), then measures of the person and of the brain both result from the same underlying cause—the state of the nervous system at the time of measurement. Because that state is captured by the neural measures, the hypothetical construct is superfluous (at least in principle, assuming you have adequate measures). Neural measurements of that state can be said to directly bring the psychological measurements into existence (i.e., to cause them), so all you need is a zero-order correlation between the two to make your point, as in Figure 1b. For example, in emotion, if psychological measures are caused by activation in some brain region, circuit, or measure of some brain state, then measures of facial muscle movement, acoustical changes in vocalizations, physical actions, and changes in autonomic measures are directly reducible to neural firing in type identity theory.

The original conception of a “hypothetical construct” also makes us more confident that such a construct is not needed in type identity theory. In 1948, as psychology was starting to struggle its way free from behaviorism, MacCorquodale and Meehl (1948) clarified the idea of a hypothetical latent construct as a process or event that is allegedly real, whose existence is inferred based on a set of observed empirical relations between measurements, but whose existence cannot be reduced to those relations. According to MacCorquodale and Meehl, a hypothetical latent construct is not merely abstracted from a set of observable measures—it relates those measures to one another by adding something (“a fictitious substance or process or idea”; p. 46). Such a latent construct has what Reichenbach (1938) called “surplus meaning”; it is a hypothesis that something exists, even if that something cannot be measured directly. The processes or events are not necessary unobservable in principle—they could be

¹In the emotion example, some older emotion models (e.g., Dewey, 1895) and some newer ones (e.g., LeDoux, 1996) ontologically reduce emotions to physical states and set aside as ontologically separate the experience of emotion as the conscious feeling of this physical state (in these models, emotion experience would be called a “homological dangler”; a mental phenomenon that does not function as a cause for any observable behavior; but see Baumeister & Masicampo, 2010).
unobservable at the moment due to temporary ignorance, a lack of sophistication in measurement or mathematical model. In fact, MacCorquodale and Meehl explicitly assume that a hypothetical construct includes inner (by which they mean neural) events that will someday be discovered (e.g., 1948, pp. 105–106). Once the inner (neural) events are specified, a theoretical construct with surplus meaning is no longer necessary to translate type identity theory into a measurement model. The “underlying attribute,” as Kievit et al. call it, is the state of the nervous system.2

In describing identity theory as an effect indicator model with a hypothetical latent construct as in Figure 1a, Kievit et al. assume that psychological and neural measurements are on “equal footing” because both imperfectly reflect the true state of the underlying attribute. But if we really assume that mental states can be ontologically reduced to brain states, as type identity theory does, then the two sorts of measurements are not on equal footing. Any measure of the person is dependent on the conditions of the brain, and so any measure of the brain will have causal ascendancy. (In a certain sense, this has to be correct—unless you are a dualist, psychological measurements, in the end, will have to be causally reduced to the brain. This does not mean that type identity theory is correct, however. The fact that mental states are caused by brain states cause mental states does not mean that one should be ontologically reduced to (i.e., merely defined as nothing but the other; I return to this issue later in the commentary.)

Figure 1b is a very simple measurement model and would need to be expanded to include multiple measures. Kievit et al. drew an expanded version of an effect indicator model that involves a hypothetical latent construct (Figure 2a), again where the latent construct is a hypothetical mental faculty (their example is intelligence). Using a similar logic to that previously laid out, however, I would draw the measurement model for type identity theory as in Figure 2b. Multiple measurements of a brain state combine to produce an estimate of that state, which in turn causes the psychological measurements. MacCorquodale and Meehl (1948) called this an “intervening variable” or an “abstractive concept.” The abstractive concept is estimated as a straightforward empirical summary of the measured variables that constitute it. This kind of construct appears in structural equation models that are called “causal indicator” or “formative” models (Bollen & Lenox, 1991). Here, neural measures are not expected to correlate with one another because they add together in a linear fashion and this aggregate realizes or constitutes the latent construct in question. Adequate measurement of the construct is dependent on measuring the correct variables. Each measure is expected to contribute unique variance to the construct, so that any small variation (again, not due to measurement error) that occurs in the brain-based measures will produce a real change in the latent construct itself (because the latent construct supervenes on the neural measures that constitute it). We might be tempted to add in a latent construct for the psychological state, as in Figure 2c, but eliminative identity theory (e.g., Feyerabend, 1963) would have us believe that the psychological can be ontologically reduced to

---

2Frankly, from the standpoint of identity theory, it is not clear what the “surplus meaning” would actually refer to. What is the nature of the nonobserved events or processes that are causing both the neural and the psychological measurements?
the physical, and thus the psychological construct is superfluous.

The philosophical model represented in Figure 2b is elegant and intuitive, and it frames a hypothesis that psychology has been wrestling with for over a century. At this point, however, it is possible to marshal a lot of empirical evidence to show that it is not correct. There is no one single brain region, network, or broadly distributed brain state for intelligence, or for memory, or even for any type of emotion. Take, for example, the category “fear.” There are well-articulated brain circuits for the behavioral adaptation of freezing, for potentiated startle, and for behavioral avoidance (Davis, 1992; Fanselow & Poulos, 2005; Fendt & Fanselow, 1999; Kopchia, Altman, & Commissaris, 1992; LeDoux, 2007; Vazdarjanova & McGaugh, 1998). These three circuits are distinct from one another, but none of them count as the brain circuit for the category “fear.” This means that measurement model in Figure 2b is empirically false when the latent construct (i.e., the brain state) is assumed to reflect a mental faculty or “type.” Faculty psychology is dead and should be given a respectful burial.

But what if the measurement model in Figure 2b was generalized to depict individual instances of a psychological category rather than the category as an entity, as in Figure 3a? That is, what if the latent construct referred to a mental state, rather than a type of mental faculty. In Figure 3a, each set of neural measurements constitutes an abstractive latent construct corresponding to a different brain state, which then produces a set of psychological measurements for a particular mental instance. These various brain states could all be variations within the same abstract psychological category. This is the token version of identity theory.

The token version of identity theory gives no ontological power to abstract psychological categories. Instead, it assumes that individual instances of a category (“tokens”) are equivalent to individual states of the nervous system (e.g., Davidson, 1980; Fodor, 1974; Taylor, 1967). Continuing with the emotion example, this is illustrated in Figure 3b. For example, token identity theories of emotion (e.g., James, 1890) assume that an individual instance of, say, fear, occurring in an individual person in a particular context is identical to a distinctive physical state, so that, for example, this instance of fear will be realized as its own specific state. According James, the perception of this state is the emotion (by which he meant the experience of emotion). Other specific instances of fear in that person, or in another person, would correspond to different distinctive physical states (depicted in Figure 3c). In token identity theory, correspondence between mind and brain is thought to reside at the level of the mental instance, and abstract categories are assumed to be folk categories with limited use for scientific induction (e.g., James’s psychologist’s fallacy; James, 1890). As in type identity theory, the correlation between the neuronal and psychological measures is thought to reflect the fact that one state (the mental state) is literally identical to the other state (the neuronal state), but in token identity theory, each brain/mental state is an instance of more abstract psychological category. There is no assumption that the variety of instances belonging to the same abstract psychological category share a physical substrate that makes them a member of that, and only that, category (James made this very point about emotion categories like anger and fear). This is why abstract psychological categories are not suitable to support scientific induction about mechanisms (although they might do a perfectly fine job at describing a phenomenon; cf. Barrett, 2009a).
Is the Measurement Model for Token Identity Theory at Odds With the Model for Supervenience Theory?

Although Kievit et al. present identity and supervenience theory as two philosophical traditions with distinct measurement models, the difference between them is less striking when using token identity theory in the comparison. There are many subtle flavors of supervenience theory (e.g., Chalmers, 1996; McLaughin, 1997; Seager, 1991; Searle, 1992), but the central idea of this philosophical position as a perspective on mind:brain correspondence is that a given kind of mental category can be realized by multiple states of the nervous system (Putnam, 1975), as represented in Figure 4. If two mental states differ in their content, then they must also be physically different—this seems obvious. But the reverse is not true—two mental states can belong to the same type of category (a mental state at two instances within the same person, or a mental state across different people), and yet their corresponding brain states can be different. (Using the concept of Neural Darwinism [Edelman, 1987], it is possible that even the same token mental state can be realized by multiple brain states within a person, but that is beyond the scope of this commentary.) So in supervenience theory, there can be no psychological difference without a brain difference, but psychological sameness does not imply neural sameness. For example, several recent publications have offered supervenient models of emotion (Barrett, 2000, 2006b; Coan, 2010; Russell, 2003). By comparing Figure 3a (representing token identity theory) and Figure 4 (representing supervenience theory), it is clear that the two are not really inconsistent with one another. The main difference is that in token identity theory (Figure 3a), a latent psychological construct is philosophically redundant with the latent physical construct. In supervenience theory (Figure 4), the psychological construct is not superfluous. It exists and, as we will discuss shortly, it is real in a particular way, but one cannot use backward inference to infer the exact cause of something (the brain state) from its product (the mental state).

Supervenience is consistent with the idea that neurons can be functionally selective for a psychological event in a given instance, even if they are not functionally specific to that psychological event. For example, we might observe consistent activation of the amygdala in the perception of fear (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, in press) and in animals who learn to anticipate a shock when presented with a tone (LeDoux, 1996), but this does not mean that we can infer the existence of fear when we observe an increase in amygdala activity (for a recent discussion, see Suvak & Barrett, 2011).

Although there are many varieties of supervenience theory, here we are concerned with two: constitutive and causal (Searle, 1992; similar to Chalmers's distinction between logical and natural supervenience or Seager's distinction between constitutive and correlational supervenience). The difference between constitutive and causal supervenience lies in the nature of the latent psychological state—is it elemental or emergent?

Constitutive supervenience is typically used to show that higher level properties (of the mind) can be directly derived from the properties or features of the lower level causes (in the brain). Facts about the mind merely redescribe facts about the brain, so that the latent psychological construct in Figure 4 is abstractive—it is a psychological label that names a state that merely intervenes between the brain state and the behavioral measures. So from the perspective of constitutive supervenience theory, Figure 4 depicts the assumption that neural measurements constitute the brain state that can then be redescribed as a psychological state that is measured by a set of person-level variables.

Causal supervenience is the more typical version of supervenience theory and is used to show how higher level properties of the mind depend on lower level properties of the brain without being reduced to them (i.e., to show how higher level constructs can be causally reduced to these lower level constructs without ontologically reduced or being merely defined in terms of them). In causal supervenience, the latent psychological construct in Figure 4 would be hypothetical in nature—a psychological state that is more than the sum of its parts (this is easier to see in Figure 7, which is an elaboration of Figure 4). Some kind of law is required to get from the physical (neural measurements) to the mental (psychological measurements), and here the concept of emergence is usually invoked. Emergence typically arises from a complex system where the collective behavior of a large assembly of more
COMMENTARIES

simple elements produces system-level properties. These novel properties are irreducible (they are distinct in existence from the more basic elements that give rise to them) and unpredictable (not due to temporary ignorance but because the starting values, context, and the interactions between more basic elements produce probabilistic outcomes). Emergent properties are also conceptually novel, in the sense that the assembly of more basic elements can be described effectively only by introducing a new concept that is ontologically new with respect to the more basic elements (i.e., the concept does not exist at the lower level of the elements). Sometimes, but not always, the emergent phenomenon must have causal powers that the lower level elements do not (this issue is typically discussed as the question of downward causation; Campbell, 1974; Kim, 1999).

Many psychological phenomena have been described as emergent (McClelland, 2010), including emotion (e.g., Barrett, 2006b; Clore & Ortony, 2008; Coan, 2010).

It is not clear whether Kievit et al. had constitutive or emergent psychological constructs in mind when they formalized supervenience theory as neural measures instantiating a latent psychological construct within a causal indicator model (their depiction is represented here in Figure 5). This lack of clarity occurs because it is not possible to empirically distinguish between the two versions of supervenience theory using the mathematics of a causal indicator measurement model (in either Figures 4 or 5). In principle, a latent psychological construct can be determined by the constellation of neural measurements either directly in an additive sense (as in an abstractive construct) or via some transformation in the emergent sense (as in a hypothetical construct). In practice, however, the latent constructs in a causal indicator measurement model must be, necessarily, only abstractive; this is because, mathematically, the construct is realized by a linear combination of its elements. The math does not yet exist to estimate an emergent phenomenon using structural equation modeling. An emergent property is a system-level property that is dependent on the organization of the system’s elements or parts, and there is nothing in a causal indicator measurement model that allows us to model the configurations that produce emergence (not even interactions between the constructs). So a causal indicator model can be used to depict emergence in a heuristic sense (e.g., Barrett, 2000, 2006a; Coan, 2010), but not in an actual mathematical sense.

If the math does not exist to model emergent phenomena within traditional psychometric approaches, then this might present a limitation for using those approaches to formalizing philosophy as measurement models. This would be a real scientific setback, because it means we don’t have the statistical tools we need to test our theoretical ideas. I suspect that in the end, new mathematical formulations will be necessary to properly map the brain to the mind. For the present, however, it makes sense to forge ahead with the mathematical tools that are available to most psychologists. Perhaps we can better formalize the philosophy of mind:brain correspondence by using a different set of psychological constructs that allow for a more tractable and testable model and that can be adapted later to better reflect the theoretical idea of emergence.

What Are The Best Categories for Use in Bridging the Gap Between Mind and Brain?

To try to understand how the brain creates the mind (and therefore how measurements of the brain are linked to measurements of the person), we have already established that a good first step is to discard abstract psychological categories of mental faculties for scientific use (at least in the fields of neuroscience and psychology). The earliest psychological scientists, like Wundt and James, advocated this move on logical grounds, and as noted previously in the discussion on type identity theory, a century’s worth of research now supports it. Such a move is consistent with supervenience theory, where the mental supervenes on the physical, so that multiple physical configurations can produce instances of the same psychological category (Figure 4). Emotion and cognition do not exist as natural kinds of psychological causes, nor can it be said that their interaction causes behavior (Duncan...
& Barrett, 2007). For psychology, such categories do not allow us to accumulate knowledge about how their instances are caused. Neither do they allow the most precise scientific predictions.

Yet a science of mind:brain correspondence that focuses on tokens (instances of mental states) does also not allow for much scientific induction or prediction either. Psychologists know this, because we have been down this road before. This was one of the lessons of behaviorism. In the formative years of scientific psychology (in the late 1800s), when psychologists first realized that constructs for mental faculties like emotion, cognition, perception, intelligence, attention, and so on, could not do the work of psychological science, they moved toward a science of instances, which led them to functionalism. To have a science of instances, they tried to discover something about the specific contexts in which each instance occurred, its specific causes, and its specific effects or the outputs that derived from that instance. This is a very inefficient way of doing science, because the fine-grained descriptions would be very complicated and unrealistic to generate in any comprehensive way (Dennett, 1991).

To this inefficiency on the psychological end of the equation add also the ambiguity about the level of measurement that is needed to adequately measure the brain. Do we measure the activation of single neurons, of columns or groups of cells, of anatomical regions, networks of regions, or (like Kievit et al.) very molar measurements like brain volume? The prospect of increasing and unrelenting complexity becomes quickly overwhelming, and there appears to be no principled level of measurement that can be proclaimed (the measure usually depends on the goals and proclivities of the scientist; Barrett, 2009a; Dennett, 1991).

Historically, a science of instances also turned out to be a perilous path for psychology, because to make a science of instances more tractable, the mind was essentially defined out of existence. Behaviorism was only a stone’s throw from functionalism. Mental states were ontologically reduced to physical instances that could then be more easily measured and catalogued. As a result, a science of instances produced a false kind of psychology, because both a functionalist and a behaviorist approach to behavior failed to capture much about the mind that is scientifically useful. The mind cannot be simply reduced to easily measurable causes and effects. Its contents also have to be described. Many philosophers have made this point (but for a particularly good discussion, see Searle, 1992). For this reason, a psychological level of description is needed. To properly describe the brain’s function, we must translate it into human terms. Neurons need to be understood not only as collections of cells but also for their functions within a person’s life. A neural process becomes a mental process when it plays a role in the organism (Lewes, 1875). As Kievit et al. correctly point out, the identification of mental constructs, even if they are neurologically grounded, depends on the psychological part of the model.

So, we find ourselves in the interesting position of needing supervenience to derive the mental from the physical but also of suspecting any version of supervenience that involves psychology constructs for mental faculties of the sort that populate folk psychology. We require a translation from brain to mind that at once preserves some aspects of token identity theory and that also addresses the need for a level of description at the psychological level that does not resort to faculty psychology.

One proposed solution is a model where (1) combinations of activations in various neuronal groups combine to realize the activation of a distributed brain network that itself is redescribed in the most basic psychological terms (i.e., it is psychologically primitive), such that (2) the interplay amongst these basic ingredients realize emergent mental states (Barrett, 2009a). This perspective, called psychological construction, hypothesizes that all mental events can be reduced to a common set of basic psychological ingredients that combine to make instances of many different psychological faculties. So, unlike a faculty psychology approach, which assumes there are “cognitive” processes that produce cognitions (e.g., a memory system that produces memories), “emotional” processes that produce emotions (e.g., a fear system that produces fear), and “perceptual” processes (e.g., a visual system that produces vision), psychological construction hypothesizes that psychological primitives are the basic elements of the mind that combine to make instances of cognition, emotion, and perception (and so on). Psychological construction models, at least in a nascent state, stretch back to the beginning of psychology (e.g., for a review see Gendron & Barrett, 2009), but because they are largely unintuitive, they are relatively rare as fully developed theories in psychology (or for mind:brain correspondence).

There are two varieties of psychological construction, one elemental and the other emergent, corresponding to the two varieties of supervenience theory. Elemental psychological construction models ontologically reduce mental categories to more basic psychological operations, as depicted in Figure 6a. Here, measures of the person reflect psychological primitives, and it is only in perception, after the fact and as a separate mental state, that humans categorize the measurable psychological event an instance of “fear,” “memory,” “perception,” and so on. For example, Figure 6a can be adapted to represent Russell’s (2003) model of emotion as in Figure 6b, where neurons
activate to produce a state of core affect (Posner, Russell, & Peterson, 2005) that can be measured by a person’s subjective feeling, their autonomic physiology, their facial muscle movements, and their vocal acoustics. In this model, the psychological primitive “core affect” is represented as the common cause for a variety of behavioral measures, meaning that those measures should correlate strongly with one another (and in the measurement of affect, they do tend to correlate; for a review, see Barrett, 2006a).

Emergent psychological construction is similar to its elemental cousin, except that instances of mental categories are thought to emerge from the interplay of the more basic operations that cause them, so that the resulting mental states cannot be merely defined in terms of those more basic parts (i.e., mental instances can be causally reduced to those operations but not ontologically reduced to them), as in Figure 7a. By analogy, we can think of these basic operations as combining, like ingredients in a recipe, to produce the instances of (or token) mental states that people give commonsense names to (and that correspond to instances of folk psychology) categories like “emotion” (or “anger,” “sadness,” etc.), “cognition” (or “belief,” “memory,” etc.), “perception,” “intelligence,” and the like (Barrett, 2009a). Emergence cannot be easily modeled in standard psychometric models, and so it is necessary to add a hypothetical latent construct to represent the emergent mental state as the instance of a folk psychological category that is created.

Inspired by the scope of the earliest psychological models, our lab introduced the first psychological construction approach to mind:brain correspondence that we know of and published several papers articulating the key assumptions and hypotheses of the model (Barrett, 2006b, 2009b; Barrett & Bar, 2009;
Barrett & Bliss-Moreau, 2009; Barrett & Lindquist, 2008; Barrett, Lindquist, Bliss-Moreau, et al., 2007; Barrett, Lindquist, & Gendron, 2007; Barrett, Mesquita, Ochsner, & Gross, 2007; Barrett, Ochsner, & Gross, 2007; Duncan & Barrett, 2007; Gendron & Barrett, 2009; Lindquist & Barrett, 2008a, 2008b). Our working hypothesis is that every human brain contains a number of distributed networks that correspond to the basic ingredients of emotions and other mental states (like thoughts, memories, beliefs, and perceptions).\(^3\)

\(^3\)Taking inspiration from connectionist and network approaches to the brain (e.g., Fuster, 2006; Mesulam, 1998; O’Reilly & Munakata, 2000; Poldrack, Halchenko, & Hanson, 2009; Raichle & Snyder, 2007; Seeley et al., 2007; Smith et al. 2009), we hypothesized that basic psychological ingredients correspond to distributed functional network of brain regions. Like ingredients in a recipe,
In my lab’s model of emotion, for example (depicted, in part, in Figure 7b), categorization is treated as a psychological primitive—it is part of the emergence of an instance of emotion, not something that comes after the fact that stands apart as a separate mental event. Our model views folk psychology categories as having meaning, not as explanatory mechanisms in psychology, but as ontologically subjective categories they have functional distinctions for human perceivers in making mental state inferences that allow communicating about and predicting human action (for a discussion see Barrett, 2009a).

In both the elemental or emergent variety, a psychological construction approach to mind-brain correspondence is an example of what Dennett (1991) referred to as the design level of the mind. By adopting the ontology of the design level, it is possible to both support scientific induction and predict “sketchily and riskily,” as Dennett (1991, p. 199) put it. Dennett (1996) recommended that a good starting point for the ontology is the intentional stance of a human perceiver. In fact, psychology did begin with the intentional stance, and it led us astray for almost a century by having us mistakenly treat folk faculty psychology concepts as scientifically causal. If psychological states are constructed, emergent phenomena, then they will not reveal their more primitive elements, any more than a loaf of bread reveals all the ingredients that constitute it.

My lab takes the view that this ontology is a work in progress and that it can be inductively discovered by more systematically investigating how that the same brain regions and networks show up across a variety of different psychological task domains (for a recent discussion, see Suvak & Barrett, 2011). In this way, it is possible to ask whether a brain region or network is performing a more basic process that is re-occurred across task domains (for a recent discussion see Barrett, 2009a).

One advantage of a psychological construction approach is that it captures insights from both token identity and causal supervenience theories. Psychological construction honors the idea that different levels of neural measurement can be described with different psychological functions. In fact, emergence very likely occurs at other levels of the model (e.g., as neurons configure into networks or perhaps if networks combine to make psychological primitives). This is a feature, not a bug, in psychological science. Furthermore, psychological construction acknowledges that a set of neurons can be described as having one psychological function when they participate in one brain network but another function when they participate in a second network. This is not a failure of cognitive neuroscience to localize function but rather an inherent property of brain function and organization. In addition, psychological construction avoids what Dennett (1996 called the “intolerable extremes of simple realism and simple relativism” (p. 37). It also gives the science of psychology a distinct ontological value and reason for existing in the age of neuroscience (cf. Barrett, 2009a). And whereas faculty psychology categories might be completely ontologically subjective (and made real by the collective intentionality in a group of perceivers, like members of a culture; cf. Barrett, 2009a), psychological primitives might be more objective, in the sense that they correspond to brain networks that are “out there” to be detected (although perhaps not anatomically). That being said, it cannot be said that these networks are

for psychological primitives comes from an inductive analysis of a meta-analytic database for neuroimaging studies of emotion (Kober et al., 2008). Using cluster analysis and multiple dimensional scaling, we identified six functional groupings consistently coactivated across neuroimaging studies of emotional experience and emotion perception. These functional groupings appear to be task-related combinations of distributed networks that exist within the intrinsic connectivity of the human brain. Intrinsic connectivity reveals many topographically distinct networks that appear to have distinct mechanistic functions (Corbetta & Shulman, 2002; Corbetta, Patel, & Shulman, 2008; Dosenbach et al., 2007; Seeley et al., 2007; Smith et al., 2009; Sridharan, Levitin, & Menon, 2008), some of which appear similar to the psychological ingredients we proposed in our original psychological construction models (e.g., see Barrett, 2006b, 2009a).

4 Complex psychological categories like emotion, cognition, perception, intelligence, and so on, are not real in an objective sort of way—they derive their reality from the intentionality that is shared by a group of people (i.e., they are folk categories that are ontologically subjective; for a discussion, see Barrett, 2009a). Therefore might retain their scientific use as descriptions of mental states that require explanation, or in sociology and other social sciences that occupy a different positions in the ontological hierarchy of sciences, but they do not themselves correspond to mental mechanisms.

5 Intrinsic connectivity networks are identified by examining correlations in low-frequency signals in fMRI data recorded when there is no external stimulus or task (hence this misnomer “resting state” or “default” activity; Beckmann, DeLuca, Devlin, & Smith, 2005; Biswal, Yetkin, Haughton, & Hyde, 1995; Buckner & Vincent, 2007; Greicius, Krasnow, Reiss, & Menon, 2003; Fox et al., 2005). The temporal dynamics of these low-frequency signals reveals networks of regions that increase and decrease in their activity together in a correlated fashion.
the “truest” level of brain organization in a perceiverto-independent kind of way. They are not entirely independent of the goals and needs of human perceivers because they are the psychological categories that we find most useful (see Dennett, 1996, p. 39; e.g., Wilson, Gaffan, Browning, & Baxter, 2010). In this way, it is possible to discuss causal reduction without being a “neophrrenologist” and entering into a kind of ontological reduction that is not tenable. Thoughts and feelings do not exist separately from the neurons that create them in the moment. But they cannot necessarily be reduced to the firing of those neurons for our purposes, either for ontological (emergent) or practical (design) reasons.

Conclusions

It has been said that although physics, chemistry, and biology might be the hard sciences, psychology is the hardest science of all, because in psychology we must make inferences about the mental from measures of the physical. Anything that helps scientists to be explicit about their philosophical assumptions in making these inferences, and forces them to be clearer about manner and viability of testing their own hypotheses, is a valuable tool. In this regard, Kievit et al.’s article was an inspiration to take the ideas of psychological construction and attempt to formalize them as measurement models. This attempt will hopefully increase the likelihood that experiments will be conducted that can properly test those models. For example, the majority of studies that have been published on emotion thus far do not involve anything other than the most typical instances of emotion and so cannot be used to properly test the basic tenants of psychological construction. Published research often treats emotions as elemental entities rather than as end states to be created and deconstructed. Studies relevant to psychological construction are just now just starting to be run and published (Lindquist & Barrett, 2008b; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, in press), and we hope that clear measurement models will encourage more research and future mathematical developments (for an interesting example of using intrinsic networks to explain a psychological task, see Spreng, Stevens, Chamberlain, Gilmore, & Schacter, 2010).

Kievit et al.’s article is also valuable because it reminds us that there is no “value-free” way to describe the relation between measures of the brain and measures of the mind. This observation is not specific to understanding mind:brain correspondence, of course. It is true whenever scientists take physical measurements and make psychological inferences from them. This is because psychological constructs are not real in the natural sense—they are real in subjective sense, and therefore they are subject to human goals and concerns. It is perfectly possible to have a mathematically sophisticated science of the subjectively real—just look at economics.

Still, challenges remain for Kievit et al.’s approach. Foremost is, practically speaking, the scalability of their approach. Their versions of identity and supervenience theory were tested using very molar measures (e.g., brain volume), but it is an open question how well this approach will work with other measures of the nervous system. Another challenge is the mathematics of standard psychometric theory. There is a lot we can do to test our philosophical notions with what we have, but the idea that might turn out to have the most scientific traction (emergence) seems to require a different set of statistical models to properly test it.

Perhaps more important than the practical obstacles is an emotional one that some might experience when reading Kievit et al.’s article. When we, as scientists, do not formalize our measurement models to lay bare our underlying philosophy of science, we are free to say (or write) one thing and mean another. And sometimes we do just that, not because of any mal-intent but more because we are unaware of the philosophical implications in how we are interpreting our data. When measurement models are excavated for their philosophical foundations, however, people have fewer opportunities for inconsistency. Being forced to be explicit in our assumptions, perhaps in their starkest form, allows us to realize that we might actually mean something other than what we intend. We give up our wiggle room. This can sometimes cause hard feelings, just as when a therapist acts the mirror to reveal a self-deception. One defense, in science, is to invoke the notorious “straw man” argument. But discomfort is not always a sign that someone intends harm—sometimes it is simply an indication that our own deeply held beliefs require a closer look, a little more deliberation, and even a change in point of view. What is intuitive is always more comfortable than what is not. But in science, such comfort rarely indicates that an idea or hypothesis is true.

Acknowledgments

Preparation of this article was supported by the National Institutes of Health Director’s Pioneer Award (DP1OD003312), a grant from the National Institute of Aging (AG030311), and a contract with the U.S. Army Research Institute for the Behavioral and Social Sciences (contract W91WAW-08-C-0018). The views, opinions, and/or findings contained in this article are solely those of the author(s) and should not be construed as an official Department of the Army or DOD position, policy, or decision. Thanks to the members of my Interdisciplinary Affective Science Laboratory for stimulating discussions on the link between
philosophy of mind:brain correspondence and psychological construction.

Note

Address correspondence to Lisa Feldman Barrett, Department of Psychology, Northeastern University, 125 Nightingale, 360 Huntington Avenue, Boston, MA 02115. E-mail: l.barrett@neu.edu

References


Philosophical Naturalism and Scientific Method

Brian D. Haig
Department of Psychology, University of Canterbury, Christchurch, New Zealand

...it is within science itself, and not in some prior philosophy, that reality is to be identified and described.
—Quine (1981)

Contemporary philosophy of mind is an important source of psychological insight that is frequently ignored by psychologists in their research deliberations. Conversely, scientific theories and methods are resources that many philosophers of mind deem irrelevant to their philosophical work. However, there is a well-established outlook in contemporary philosophy known as naturalism, which asserts that philosophy is continuous with science and which attempts to formulate and evaluate philosophical theories by using the research findings and investigative means of the various sciences. Naturalism, in all its variety, is probably the reigning outlook in contemporary philosophy, and it is especially popular in the philosophy of science and the philosophy of mind.

In their informative and innovative target article, Kievit and colleagues (this issue) profitably combine resources from the philosophy of mind, scientific methodology, and empirical science to demonstrate that two prominent theories of the mind–body relation—the identity theory and the supervenience theory—can be articulated and tested using structural equation modeling methods. Kievit et al. are concerned to move these two speculative theories from their customary position of metaphysical isolation in philosophy into the realm of cognitive neuroscience. In doing so, the authors have two main goals. The first is to demonstrate how two well-known theories in the philosophy of mind can be made scientific by testing them empirically. As such, this part of their project can be regarded as a contribution to naturalized philosophy of mind. The authors’ second goal is to show how the statistical methods of structural equation modeling can be employed in cognitive neuroscience to illuminate the relation between psychological and neurological properties. In the course of pursuing these goals, Kievit et al. have instructive things to say about the relation between philosophy of mind and cognitive neuroscience, and about the construction of theories and the use of scientific methods. Written by authors who combine philosophical, psychometric, and substantive psychological expertise, the target article is a testament to the idea that the disciplines of philosophy and psychology can be profitably conjoined.

In this commentary, I focus on matters that arise from the authors’ pursuit of their first goal—the naturalization of philosophy of mind—by considering a number of ways in which philosophy and scientific psychology might be brought together to their mutual advantage. My focus for the most part is on methods of inquiry. I begin with a big picture consideration by briefly discussing the philosophies of empiricism and scientific realism and their attitudes to naturalized philosophy. I then comment on Jagewon Kim and David Lewis’s views of naturalism in the philosophy of mind and consider their relation to the naturalism of the target article. Thereafter, I comment specifically on some limitations of structural equation modeling, the method of choice in the target article, and the useful but neglected method of inference to the best explanation. After that, I suggest that philosophy and psychology might be encouraged to use each others’ methods to their mutual advantage. Before concluding, I draw attention to the philosophy of normative naturalism. I suggest that it can help psychologists better understand the foundations of behavioral science methodology.

Traditional Empiricism: Philosophy and Science Separated

Given that psychology grew out of philosophy, and has operated as a self-conscious science for more than 100 years, the suggestion that psychology and philosophy should join forces to their mutual advantage will strike some psychologists as unwelcome advice. For psychology is still saddled with the traditional empiricist idea that philosophy and science are different in kind, both in subject matter and in method. Many of its practitioners think that philosophy is a discipline with its own unique problems and investigative styles based on a priori armchair reflection, whereas psychology is regarded as a science whose substantive claims are founded a posteriori on empirical evidence.

Consistent with this understanding of the differences between the two disciplines, the philosophy of standard empiricism is taken to be a privileged and unreviseable philosophy of science. It is deemed to exist prior to, and apart from, science and provide a
foundation of certain, or near certain, knowledge about
science. As an autonomous and insular discipline, this
philosophy has not looked to learn systematically from
the various sciences.¹

Consistent with this empiricist outlook, philosophy
is viewed by a majority of psychologists as a dispens-
able luxury that has little, if anything, to do with their
workaday world as scientists. Because of this attitude
to philosophy, I think it will come as both a surprise
and a puzzle to many readers that the authors of the tar-
get article are intent on evaluating the scientific worth
of philosophical theories.

Naturalistic Realism: Philosophy and Science
Conjoined

It is well known in philosophical circles that ortho-
dox empiricism is an outmoded philosophy of science
and that its conception of philosophy is difficult to de-
defend. Scientific realism is the major alternative to em-
piricist philosophy of science.² One attractive form of
this philosophy, naturalistic realism (Hooker, 1987), is
so called because it is a realist theory of science based
on naturalism. According to this theory, scientific rea-
soning, including theorizing, is a natural phenomenon
that takes its place in the world along with other natural
phenomena. Further, philosophy and science comprise
a mutually interacting and interconnected whole. As a
philosophical theory about science, naturalistic realism
has no privileged status and may be revised in the light
of scientific knowledge. Similarly, the naturalistic rea-
ist foresee that philosophical conclusions, tempered
by scientific knowledge, may force changes in science
itself.

According to one influential view of naturalism,
philosophy and science are interdependent. This inter-
dependence takes the form of mutual containment (cf.
Quine, 1969), though the containment is different for
each. Philosophy is contained by science, being located
within science as an abstract, critical endeavour that is
informed by science. Science is contained by philoso-
phy because the latter, amongst other things, provides
a normative framework for the guidance of science.

Naturalistic realism maintains that philosophy of
science is that part of science concerned with the crit-
ical in-depth examination of science in respect of its
presuppositions, aims, methods, theories, and institu-
tions. Philosophy of science naturalized is in a sense
science applied to itself: It employs the methods of
science to study science; it is, where appropriate, con-
strained by the findings of science; and it is itself a
substantive theory of science. As such, naturalized phi-
losophy of science is at once descriptive, explanatory,
advise, integrative, and reflective of science. Being
positioned within science, naturalistic philosophy is
well placed to study science, learn from science, and
instruct science. The proponents of naturalized phi-
losophy of science are many and varied (Rosenberg,
1996). Prominent among them are Richard Boyd, Clif-
ford Hooker, Ronald Giere, Larry Laudan, and Philip
Kitcher.

Not all of these philosophers are scientific realists
(and not all scientific realists are naturalists), which
raises the question, Why is it advantageous to com-
bine scientific realism and naturalism in a philosophy
of naturalistic realism? One reason is that naturalism
is the best methodology we have available to us; it gives
us our best methods and encourages us to constrain
our theorizing in light of reliable scientific knowledge.
A further reason is that its principled commitment to
both anti-anthropocentrism and fallibilism affords us
a realistic defence of realism, one that is true to our
makeup as cognizers. Finally, by embracing natural-
ism, realism becomes an integrated whole that afford us
the best explanatory theory of the cognitive dynam-
ics of science (cf. Hooker, 1987). I briefly remark on
the explanatory worth of scientific realism later.

What of realism regarding the mental? I think it is
evident that we have good reason to be realists about
mentality in both philosophy of mind and scientific
psychology. Realism in the philosophy of mind has
all the characteristics of philosophy of science operat-
ing in the domain of the mind. In both scientific and
(lay) folk psychology, the explanatory and predictive
achievements of our theories about the mental, modest
though they often are, are sufficient to warrant a realist
outlook (cf. Fletcher, 1995).

Varieties of Naturalism

There are many different forms of naturalism in phi-
losophy, and there is considerable debate about how
naturalism should be conceived (Kitcher, 1992). Mod-
eral discussions of naturalism often begin by referring
to Quine’s efforts to rehabilitate naturalism in episte-
ology (Quine, 1969), and I do the same. I then look
at the naturalist commitments of Kim and Lewis in
their respective theories of the mind/body relation, be-
fore briefly comparing them with the naturalism of the
target article. I conclude my selective overview of dif-
ferent naturalist positions by pointing to examples of

¹Despite its separatist conception of philosophy, classical em-
piricism’s prescriptions for the conduct of inquiry have exerted a
palpable influence on psychological science (e.g., the attraction of
operational definitions, the heavy use of Fisherian statistical pro-
dures, and the steadfast neglect of theory by the publication manual
guidelines of the American Psychological Association).

²The major debate between realists and empiricists in contem-
porary philosophy of science pivots around Bas van Fraassen’s (1980)
constructive empiricism and his criticisms of scientific realism. This
debate, and other debates between realists and antirealists, have been
widely ignored by psychologists.
naturalism in psychology, the philosophy of mind, and philosophical methodology.

Quine’s Naturalism

Quine is generally regarded as the most influential philosophical naturalist of the 20th century, and it comes as no surprise that the authors of the target article refer their readers to Quine’s (1969) landmark essay, “Epistemology Naturalized,” for philosophical reasons for adopting a naturalist stance in epistemology.3

The primary features of Quine’s naturalism are succinctly characterized by him as follows: “Naturalism [is] the abandonment of the goal of a first philosophy. It sees natural science as an inquiry into reality, fallible and corrigible but not answerable to any supra-scientific tribunal, and not in need of any justification beyond observation and the hypothetico-deductive method” (Quine, 1981, p. 72). Philosophy, understood as first philosophy, is different from science. It is a discipline that is methodologically prior to science and, through a priori reflection, fashions general truths that provide a foundation of justification for scientific inquiry itself. It is the “supra-tribunal” for science.

In rejecting this picture of first philosophy, and its assumption that philosophical knowledge can be obtained a priori, Quine conceives philosophy as broadly scientific, maintaining that all knowledge is a posteriori and that science is our best means of obtaining such knowledge. For Quine, philosophy and science are of a piece, and his naturalism can be regarded as scientific in the best sense of that term.

With his strong commitment to naturalism in epistemology, Quine has obvious high regard for scientific method. In this respect, he endorses the hypothetico-deductive method, claiming that it is a sufficient account of scientific method. In some tension with this claim, he also emphasizes the importance of theoretical virtues such as conservatism, modesty, and simplicity, in justifying scientific claims. However, Quine is not much interested in the details of scientific methods and the patterns of reasoning that their employment facilitates. And, apart from his account of language learning, he refrains from exploiting substantive scientific knowledge as a source of constraints on philosophical theorizing. This disregard for empirical constraint is in marked contrast to the naturalism of the target article, which identifies and justifies the method of structural equation modeling to specify and empirically test the supervenience and identity theories of the mind.

Kim’s Naturalism

Although Kim sees his supervenience theory as part of a naturalized philosophy of mind, it is a much weaker form of naturalism than that pursued by Kievit et al. in their target article. Kim is a well-known critic of Quine’s naturalized epistemology, faulting it because he thinks it is purely descriptive and has no place for the normative claims that he maintains are part and parcel of epistemology proper (Kim, 1988). However, many naturalists, including the later Quine, accept that a proper naturalist epistemology can be both descriptive and normative. A proper naturalistic epistemology needs to embrace the range of normative concerns that have traditionally motivated epistemologists but read- dress them from the vantage point of a naturalistic epistemology. For example, contemporary naturalistic epistemology is well positioned to provide constructive advice on how human cognizers can improve their epistemic situation by correcting their cognitive biases in a manner that traditional epistemology was neither able nor motivated to do (cf. Bishop & Trout, 2005). Laudan’s normative naturalism, which I briefly consider at the end of the commentary, is so called because it maintains that normative considerations are central to a naturalist conception of methodology.

Although Kim’s commitment to naturalism in the philosophy of mind allows for a normative dimension, he maintains that we must seek an explanation of the mind/body relation through natural science, not psychology or philosophy. He himself does not provide an explanation of the supervenience relation as such. Instead, he provides an abstract characterization of the relation in the mode of an analytic philosopher. This approach to philosophy in good part involves clarifying conceptual obscurities, detecting and correcting unreasonable arguments, and avoiding unwanted philosophical positions.

Kievit et al. chose to focus on Kim’s supervenience theory because it is both well known and well regarded. However, despite being the product of a first-rate philosophical mind, the substantive content of Kim’s supervenience theory is rather meagre, and it makes no genuine contact with the mental sciences. These facts are borne out in Kievit et al.’s necessarily brief characterization of his supervenience theory, a characterization that they are able to expand on by specifying the supervenience relation as a formative structural equation model.

Lewis’s Naturalism

Lewis was a systematic philosopher committed to realism, both scientific and metaphysical. He was a reductive materialist about the mind, maintaining that mental states are contingently identical to physical states, particularly neural states. It is appropriate,
therefore, that Kievit et al. focus on his identity theory as a potential contribution to reductionist cognitive neuroscience. But Lewis also developed explicit views about how philosophers should go about their business of philosophizing (Lewis, 1983; Nolan, 2005). I want to say something about Lewis on this score in order to better appreciate at a methodological level how his identity theory, as he regards it, actually compares with its empirical evaluation in the target article.

Lewis took philosophical inquiry to be a form of conceptual analysis, but it was conceptual analysis with a difference. Many philosophers used to regard conceptual analysis as the specification of the primary meaning of nontechnical terms that are of interest to philosophers. This was understood as something quite different from the construction and elucidation of synthetic theories. However, for Lewis, philosophical inquiry should, where possible, follow what has come to be called the Canberra Plan. According to this plan (cf. Braddon-Mitchell & Nola, 2009), which Lewis helped shape, the philosopher's first task is to engage in a priori analysis of the concepts and categories employed in everyday thought. Such thought is assembled in our folk theory of common sense, a theory that Lewis maintained we are entitled to believe in despite its unsystematic nature. Indeed, Lewis seemed to think that some commonsense claims are nonnegotiable because we do not have sufficient reason to believe anything else. For all that, Lewis insisted that the results of such conceptual analyses are modest and that they yield no empirical knowledge. The second task in the Canberra Plan is to invoke relevant a posteriori scientific knowledge about the basic nature of reality in order to satisfy those folk concepts and categories. This task involves seeking the referents of these concepts as suggested by our best scientific theories. And, the successful search for such referents justifies our application of the concepts in question.

The Naturalism of the Target Article

It is this second task in the Canberra Plan that makes Lewis's identity theory of mind a part of naturalistic philosophy. However, Lewis's naturalism is a more restrained form of naturalism than the thorough-going naturalism of the target article. In contrast, Kievit et al. take Lewis's identity theory, regard it as a substantive theory of the mind, and show that it can be scientifically evaluated by submitting it to empirical test. By specifying his identity theory as a structural equation model, they subject it to a regimentation that is quite different from the regimentation provided by the Canberra model, which Lewis himself used. Of course, at the same time, the target article provides an additional justification for the identity theory to that provided by Lewis himself.

Lewis was a system builder, and for him the plausibility of his identity theory stemmed from being part of the broad reflective equilibrium of his total philosophical system. However, Lewis went further than just relying on the vague notion of reflective equilibrium, for he justified the credence of his philosophical theories by appealing to their fruitfulness, that is, to their simplicity, unifying power, explanatory value, and their conservativeness. In this respect his methodology shares affinities with that of Quine.

Although Lewis's evaluation of his identity theory was broad ranging, he did not submit it to explicit systematic appraisal. By contrast, Kievit et al.'s justification of the identity theory is more local, stemming as it does from measures of empirical adequacy provided by structural equation modeling. Their justification is thereby more explicit and systematic than that provided by Lewis. However, I note shortly that one can provide a more comprehensive approach to theory evaluation than that adopted in the target article.

Speculative Theory and Empirical Constraint

In addition to constructing theories about science, naturalism also encourages philosophers to function as speculative scientists by fashioning substantive theories with a posteriori in their domains of interest. In contemporary philosophy of mind, Jerry Fodor, Daniel Dennett, Stephen Stich, Paul and Patricia Churchland, Fred Dreske, and many others, have engaged in the project of constructing naturalistic theories of the mind. It is a concern that their creative contributions are not being incorporated into psychology's mainstream efforts to understand the mind, for they rank among the most suggestive psychological theories currently on offer.

However, it also matters that some of these theoretical contributions are not heavily constrained by extant psychological knowledge. Take, for example, Paul Churchland's (1981) well-known critical evaluation of the scientific status of folk psychology. He argued that there is no good evidence for the existence of the beliefs and desires that folk psychology postulates and that we should discard folk psychology in favour of a neuro-scientific alternative. Churchland's

---

4 The Canberra Plan is so called because many of its early proponents were philosophers associated with the Australian National University in Canberra. Frank Jackson and Philip Pettit are prominent among them.

5 A sceptical philosopher might argue that much psychological knowledge is superficial and far from the truth, and that it, therefore, ought not to seriously constrain philosophical theorizing about the mind. I think there is something to be said for this view, but it needs to be argued for on a case-by-case basis with a detailed examination of the epistemic credentials of the knowledge claims in question.
case against folk psychology is multifaceted, but it is conspicuous by its absence of a concerted examination of the relevant empirical literature in psychology that speaks to the worth of folk psychology.

Fletcher (1995) has examined the scientific credibility of folk psychology at length. However, unlike Churchland, Fletcher examined the relevant empirical evidence in social psychology as well as the conceptual and theoretical issues that are germane to its credibility. Citing empirical studies of his own, and the research of others, Fletcher demonstrated that “under certain conditions, folk theories do rather well in explaining and predicting social behaviour” (p. 89). Naturalistically inclined philosophers have been reluctant to make full use of the relevant empirical literatures in their spheres of interest. Fletcher’s work is a good example of how theoretically oriented psychologists, whose stock-in-trade is empirical research, can helpfully contribute to a naturalist perspective on the mind.

Relatedly, the work of a number of people whom we regard as prominent psychological theorists cannot be properly understood unless we take them to be naturalistic philosophers as well as psychological researchers. Consider, for example, Piaget and Skinner. To understand and evaluate Piaget’s developmental psychological research we must see it as part of his genetic epistemology, whereas Skinner’s radical behaviourist psychology should be seen to contain a psychologically oriented treatment of knowledge processes in science.

I turn now to consider structural equation modeling, which is the method used by the authors of the target article to empirically evaluate the identity and supervenience theories. I then suggest that the method of inference to the best explanation might also usefully be employed to evaluate theories like these.

### Naturalism and Scientific Method

#### Structural Equation Modeling

By using structural equation modeling methods to evaluate the identity and supervenience theories, Kievit et al. adopt the hypothetico-deductive method of theory appraisal. The hypothetico-deductive method, which has long been the method of choice for the evaluation of psychological theories, is commonly characterized in minimalist terms: The researcher takes an existing hypothesis or theory and submits it to indirect test by deriving from it one or more observational predictions that are themselves directly tested. Predictions borne out by the data are taken to confirm the theory to some extent; those that do not square with the data count as disconfirming instances of the theory.

One feature of Kievit et al.’s research, which makes the hypothetico-deductive method fit for their purpose, is that it takes the two theories it tests as givens, irrespective of their origin. In using the hypothetico-deductive method, it matters not that the identity and supervenience theories of the mind/body relation were formulated in philosophers’ armchairs. All that matters is that they are stated in propositional form and can, with suitable specification, be made amenable to empirical testing.

It is a feature of structural equation modeling that it uses goodness-of-fit measures as a basis for judging the acceptability of the models it tests. Leaving aside difficulties in determining the corroborative value of these measures, it should be emphasized that goodness-of-fit is a criterion of empirical adequacy (Rodgers & Rowe, 2007) that by itself provides insufficient grounds for assessing the credibility of competing models. This limitation is a special case of the problem known as underdetermination of theories by data (better, empirical evidence), and an attractive solution to this problem is to supplement measures of empirical adequacy by appealing to the so-called superempirical or theoretical virtues such as explanatory power, fertility, and simplicity (cf. McMullin, 1983). Although the use of criteria such as these do not “close the gap” between theory and empirical evidence, they do reduce it, thereby enabling the researcher to manage this particular underdetermination problem.

#### Inference to the Best Explanation

As just noted, the orthodox hypothetico-deductive method takes predictive accuracy as the sole criterion of theory goodness. However, when explanatory criteria are invoked, a quite different approach to theory appraisal is employed—an approach known as *inference to the best explanation* (Haig, 2009; Lipton, 2004; Thagard, 1992). I think that this alternative perspective on theory appraisal could be used with profit to evaluate metaphysical theories like those considered in the target article. With inference to the best explanation, the ideas of explanation and evidence come together, and explanatory reasoning becomes the basis for evaluating theories: The explanatory goodness of theories counts in their favour; conversely, the explanatory failings of theories detract from their credibility.

According to Thagard (1992), inference to the best explanation is essentially a matter of establishing relations of explanatory coherence between propositions within a theory. On this account of inference to the best explanation, to infer that a theory is the best explanation is to judge it as more explanatorily coherent than its rivals.

---

6 Despite the sophistication of structural equation modeling, a number of authors have raised doubts about its use of fit indices in model selection. Barrett (2007) and McDonald (2010) are two recent expressions of concern about the difficulties in determining the fit of structural models to data.
than its rivals. Structural equation models are networks of propositions, and theories depicted as networks of propositions lend themselves naturally to evaluation in terms of considerations of explanatory coherence.

In research practice, the hypothetico-deductive method is sometimes combined with the use of supplementary evaluative criteria. When this happens, and one or more of the criteria have to do with explanation, the combined approach can appropriately be regarded as a version of inference to the best explanation, rather than just an augmented account of the hypothetico-deductive method. This is because the central characteristic of the hypothetico-deductive method is a relationship of logical entailment between theory and evidence, whereas with inference to the best explanation the relationship is one of explanation. The hybrid version of inference to the best explanation being noted here will allow the researcher to say that a good explanatory theory will rate well on the explanatory criteria and at the same time boast a measure of predictive success. Most methodologists and scientists will agree that an explanatory theory that also makes accurate predictions will be a better theory for doing so.

Although the use of structural equation modeling in psychology often involves testing models in hypothetico-deductive fashion, it also contains a minority practice that provides an example of inference to the best explanation in the sense just noted. This latter practice involves the explicit comparison of models or theories in which an assessment of their goodness-of-fit to the empirical evidence is combined with a weighting of the fit statistics in terms of parsimony indices (Kaplan, 2000). Here goodness-of-fit provides information about the empirical adequacy of the model, whereas parsimony functions as a criterion having to do with the explanatory value of the model. Both are used in judgments of model goodness. Markus, Hawes, and Thasites (2008) recently suggested that in structural equation modeling, model fit can be combined with model parsimony, understood as explanatory power, to provide an operationalized account of inference to the best explanation. They discussed the prospects of using structural equation modeling in this way to evaluate the comparative merits of two- and three-factor models of psychopathy. It would be both interesting and informative to see structural equation modeling used in this manner to evaluate the identity and supervenience of theories.

Sharing Philosophical and Scientific Methods

Combining philosophy and science in the manner suggested by naturalistic realism has the important implication that philosophers and scientists should use each other’s methodologies to further their work. In what follows, I briefly consider three quite different methods that have been, or might be, used in both philosophy and science.

Conceptual Analysis

The analysis of concepts has been a stock-in-trade of traditional philosophy. Despite spirited criticisms of the worth of conceptual analysis to philosophy, its clear importance in the Canberra Plan discussed earlier suggests that conceptual analysis in some form still has a useful role in naturalistic epistemology. Moreover, the utility of conceptual analysis extends to scientific practice itself. Kievit et al. mention the mereological fallacy as an example. This fallacy involves the mistake of attributing to a part what can only properly be attributed to a whole. The authors cite an example from Bennett and Hacker (2003), who maintained that claims in contemporary neuroscience like “the frontal lobe engages in executive functioning” attribute to a part of the brain what can only properly be attributed to human beings.7

The importance of conceptual analysis in helping to improve the quality of psychological science has received little attention within the discipline. However, Rozeboom (1977) has strongly urged psychologists to engage in professional critical analysis of their concepts in order to improve their thinking about substantive issues. He sees this practice as an important and neglected aspect of scientific methodology that deserves to be systematically taught and practiced along with statistical methodology. Rozeboom understands conceptual analysis in rather broad terms to include clarifying the meaning of terms, identifying the depth-grammar of concepts, probing the ideational structure of theories, and evaluating the quality of scientific reasoning. This broad undertaking he called metathink, which is a detailed working out of the two pragmatic questions. “What do I mean?” and “How do I know?” In short, conceptual analysis is a neglected, but useful, addition to the psychological researcher’s methodological armamentarium.

Inference to the Best Explanation Again

Earlier, it was noted that inference to the best explanation is an important approach to the justification of scientific theories. However, inference to the best explanation has also been used as a means of justifying theories in philosophy of science, metaphysics, and other branches of philosophy. For example, a number of philosophers have offered a defence of

---

7I don’t mean to suggest that we should subscribe to Bennett and Hacker’s Wittgensteinian view that empirical considerations do not bear on the process of conceptual analysis. Sytsma (2010) is a recent discussion and demonstration of the relevance of empirical investigations to conceptual analysis.
scientific realism as the best explanation for the success of science, whereas others in metaphysics have employed inference to the best explanation to justify theories about the existence of properties or universals. It should be acknowledged that the use of inference to the best explanation has been the subject of considerable debate, and there are differences between scientists and philosophers in their use of the method. Nonetheless, the use of inference to the best explanation in both science and philosophy is striking evidence of a strong methodological continuity between science and philosophy.

It is perhaps worth mentioning here that improvements in our understanding of inference to the best explanation can be had by combining insights from both philosophy and science. For example, Thagard’s (1992) theory of explanatory coherence mentioned earlier is a detailed working out of inference to the best explanation that draws from epistemology, philosophy of science, and cognitive science, including computer science. The codification of this method is made possible only by the strong integration of philosophical and scientific insights.

**Other Statistical Methods**

Kievit et al. have demonstrated how one can make use of sophisticated statistical methods that are widely used in the behavioral sciences in order to empirically evaluate philosophical theories. As they note, the quantification of the identity and supervenience theories that comes with their statistical specification as structural equation models marks a significant scientific advancement of those theories. It should be emphasized, therefore, that this practice deserves to be implemented on a larger scale.

In this regard, it is worth noting Faust and Meehl’s innovative plea for studying scientific theorizing empirically (Faust & Meehl, 2002; Meehl, 1992). These authors have vigorously argued for an improvement in the use of quantitative methods for studying the impact of historical evidence on our understanding of scientific episodes and scientific reasoning in the history and philosophy of science. This pursuit, which Meehl has dubbed *cliometric metatheory*, uses actuarial methods to supplement case study information in the history of science to better understand scientific processes. Faust and Meehl maintain that psychologists, with their psychometric knowledge and skills, are well positioned to lead developments in this discipline, which in recognition of its central use of powerful statistical methods, they call *meta-science* rather than *philosophy of science*.

Cliometric methodology comports well with Laudan’s (1996) well-known normative naturalist position in the philosophy of science. Laudan maintains that validating scientific rules should involve the use of method-based empirical information to ascertain the frequency with which particular methods are likely to promote their appropriate epistemic goals.

There is no doubt that philosophy of science can benefit from the use of scientific methods to help validate its own theories. On the other hand, contemporary philosophical methodology has considerable resources for helping us better understand and advance our knowledge of effective inquiry procedures in psychology and the other sciences. What better way to learn how to do good science than to critically examine how good scientists have actually gone about their work? Useful philosophical progress has been made in understanding the nature of research problems, the nature of experimentation, the theory of data analysis, methods of theory construction, and so on. Insights from these philosophical sources sorely need to be combined with what is good in psychology’s indigenous and neighbouring methodological practices so that psychologists can strengthen their hand in the procedural domain.

Philosophy of science also recommends itself as an excellent medium through which to convey a decent understanding of psychological theory and method. In its contemporary reconstructions of scientific research, philosophy of science explicitly identifies and critically highlights many features of science that are important pedagogically. These include frameworks, idealizations, models, unifying theories, inquiry strategies, and methodological judgments. Harré (1983) has shown how the philosophy of science can be construed as science criticism to revise and improve the pedagogy of social psychological research.

**Normative Naturalism**

My primary focus in this commentary has been on a number of methodological aspects of the broad project of philosophical naturalism. Before concluding, I want to note one more variant of philosophical naturalism—the important development in philosophy of science known as *normative naturalism*. Normative naturalism is a position, developed most notably by Laudan (1996), on how one should understand the challenging problem of the justification of scientific method. Although psychologists are much concerned with method
in their research deliberations, and behavioral science methodologists sometimes examine the performance characteristics of their methods, neither show much awareness of the role that philosophical contributions can make to an overall understanding of those methods. Laudan’s normative naturalism affords psychologists an instructive philosophical perspective on scientific method (Capaldi & Proctor, 2000).  

Laudan’s normative naturalism is a form of methodological naturalism. It is normative both because it is concerned with the nature of justification in science and because it is intended to be a source of recommendations for scientists to follow. It is naturalistic because it maintains that scientific methodology is an empirical discipline, and that as such it is part and parcel of natural science. Laudan focuses on methodological rules. For him they are to be understood as hypothetical imperatives of the form, “If you want to reach goal X, then use strategy or method Y.” To cite one of his examples, “If one wants to develop theories that are very risky, then one ought to avoid ad hoc hypotheses” (Laudan, 1986, p. 24). In similar fashion, one might formulate a more contextually specific methodological rule from the target article as follows: “If you want to empirically test the identity theory of mind, then one ought to employ a formative structural equation model.” Normative naturalism is naturalistic in that it regards methodology as continuous with scientific theories about how the world is constituted. In effect, methodology has the status of a broad empirical theory about the conduct of inquiry. Thus, methodological rules are subject to evaluation by empirical means and may be revised, or even replaced, in the light of empirical research.

In addition to formulating a position about the nature of methodology (meta-methodology), Laudan, in collaboration with others (Donovan, Laudan, & Laudan, 1988; Laudan et al., 1986), initiated a major research programme that provides case study evidence for the role that a number of well known methodological rules have played in episodes of scientific change. It should be noted here that although Laudan rejects the justification of empirical rules on a priori grounds, his insistence on obtaining empirical evidence in order to justify methodological rules does not exclude employing conceptual considerations as well. In his earlier philosophical work, Laudan (1977) stressed the importance of identifying and solving conceptual problems as an important part of theory appraisal. For him, the conceptual well-foundedness of theories requires

the scientist to identify and remove logical inconsistencies, conceptual ambiguities, and incompatibilities with established theories and methodologies. In similar fashion, I have suggested that complementary “nonempirical” criteria might profitably be employed when using structural equation modeling to evaluate statistical models.

**Conclusion**

I have written this commentary in the belief that a thorough-going naturalism should be adopted within both philosophy and psychology. Michael Devitt (2010), who is both a scientific realist and a naturalist, recently averred that “naturalism is worth dying for” (p. 180). To be an antinaturalist in philosophy, and thereby ignore science, is to deprive ourselves of the best knowledge we have about the world. To be a philosophical naturalist in science enables us to take advantage of the best that philosophical imagination and criticism can offer. In their target article, Kievit et al. have done psychology an important service by showing how one can use the naturalism inherent in scientific methods to improve our understanding of the credentials of two suggestive theories bequeathed us by naturalist metaphysics. They have also provided naturalist philosophers with an instructive example of what the use of scientific method in philosophy looks like when one gets down to tin tacks. It is to be hoped that more psychologists and philosophers will work together to develop and evaluate theories in the manner of the target article.

**Note**

Address correspondence to Brian D. Haig, Department of Psychology, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. E-mail: brian.haig@canterbury.ac.nz

**References**


Reductionism and Practicality

Edward Vul

Department of Psychology, University of California, San Diego, La Jolla, California

All models are wrong, but some are useful. — George Box (1979)

What is the best level at which to describe human cognition? We could describe it using mathematical formalisms (like Bayesian statistics) at Marr’s (1982) computational level by specifying the sources of information in the world and our own inductive biases that we draw on to make inferences and choose actions in the world. We could describe human cognition at the algorithmic level using the language of computer science, describing how people represent data, and what procedures operate over these representations to make the required computations. We could instead adopt the language of electrical engineering and talk about the neural signals, systems, and circuits that are the physiological instantiations of the algorithmic description. We could reduce further to the level of biochemistry, where we describe the individual neurotransmitters, ion channels, and chemical gradients that allow neurons to pass information between one another and generate action potentials. Of course, we needn’t stop there, as those individual neurotransmitter molecules and ions comprise atoms and subatomic particles.

So, how do we decide at which level of abstraction to operate? Because models at higher and lower orders of abstractions are all likely to be wrong, we can only answer this question practically, by hoping that some of these models are useful for predicting or manipulating some target phenomenon. When approached from such an engineering perspective, there are specific costs and benefits to operating at each level. Even in physics—a model of reductionist success—when dealing with a higher order abstraction (like classical mechanics), we will fail to account for some subtleties that would be captured at a finer scale (quantum interactions), which could end up playing an important role in the phenomenon of interest. When dealing with lower abstractions (such as particle physics), we face a vast computational challenge when trying to describe higher order phenomena (like how a ball will bounce). Thus, physical models at different levels of abstraction will prove to be more or less useful depending on the phenomenon of interest, so different abstractions are emphasized in astrophysics, mechanical engineering, electrical engineering, and quantum computing.

These trade-offs apply to predicting human cognition. If we want to predict whether a given drug will increase dopamine in Parkinson’s patients, psychological and cognitive neuroscience descriptions are practically useless: Our question is about biochemical interactions, so biochemical descriptions of the brain provide the most useful basis for psychopharmacology. In contrast, if we want to tell a neurosurgeon where to cut to avoid damaging the patient’s capacity for speech, biochemical and psycholinguistic descriptions are useless; however, theories and data from cognitive neuroscience, indicating which parts of the brain are more involved in speech production and comprehension provide the most relevant abstraction and can fruitfully guide surgery. If, however, we focus on a complex human behavior, for instance, to find the best teaching schedule in a classroom, we derive this prediction neither from the biochemical processes underlying long-term potentiation and long-term depression nor from our cognitive neuroscience descriptions of hippocampo-cortical storage loops; instead, psychological accounts of forgetting curves, testing, and spacing effects yield a powerful basis for prediction.

Connecting these different levels of description is a necessary and fruitful research enterprise. We are reassured of our scientific models at higher levels of abstraction (like trichromacy—the theory that human color vision is three-dimensional) when those models may be derived from properties at lower orders of abstraction (the existence of three cone types). Similarly, we are reassured that we are measuring relevant properties of complexly interacting elements (like receptive field size of V1 cells) when those properties can be simplified to abstractions about the important behaviors of the system as a whole (cortical magnification and the falloff of acuity with eccentricity). Thus, connecting levels of description validates models at both levels of abstraction, so there is a scientific demand for a single unified model of human behavior, cognition, and neuroscience by reducing cognitive theories to their biological underpinnings. However, there is no reason to expect that even when the levels of abstraction are united through reductionism that one level of description will emerge as the most fundamental, useful, or practical.

In the target article, Kievit et al. (this issue) describe a psychometric approach to connecting cognitive
neuroscience and psychological levels of description based on the premise that reductionism may be achieved by constructing joint measurement models of neural and psychological variables to determine the causal relationships between these variables. This seems like a particularly fruitful approach for testing particular reductionist theories—insofar as fluctuations in a cognitive variable can be well predicted by a linear weighting of fluctuations in a neural variable, one has strong evidence that researchers are looking at the correct variables. Moreover, the psychometric approach described in this article can provide a fruitful way to adjudicate which variables, from which level of description, are most effective at predicting phenomena of interest.

Nevertheless, the central challenge of reductionism is in finding the right variables at each level of abstraction, not in specifying statistical models to compare these variables. To illustrate this point, Kievit et al. show that high-level variables like intelligence and personality are not well predicted by coarse neural measures like gray/white matter volume and density in large regions of interest. But this is not surprising—few researchers would suggest that intelligence or personality amounts to the mass of one or another type of neural tissue. Moreover, there is no reason to suspect that “intelligence” or “personality” are particularly fundamental high-level variables: They are aggregate metrics of behavior.

The failure to find that coarse neural metrics can predict coarse behavioral metrics only highlights the central challenge of reductionism: Before we can reduce cognition to neural variables, we must develop theories of cognition and theories of neural computation. Only once we have both an adequate theory of cognition that can predict—rather than retrospectively describe—human behavior and a theory of neural computation that can predict how assemblies of neurons, glia, and capillaries interact will it be possible to establish meaningful reductions between variables that emerge from theories at the two levels of abstractions. Until then, reductionism will at best have only a superficial sheen of success, rather than revealing fundamental understanding that can drive practical predictions and interventions.

Note

Address correspondence to Edward Vul, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: evul@ucsd.edu

References


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 summarize 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 explanations 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 is thus obvious.
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, & Augelmann, 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...
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 lambda21 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., 2011).

---

\(^1\) 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 purpose 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 construct 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

2 Although 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.

3 Haig (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 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 representations 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 involves 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. 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

---

4 Although 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 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, Dennissen, & 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., Baggozzi’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
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).
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 philosophical 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 can we 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
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 Y 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 parsimonious (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 models 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, Nesselroade, & 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


Kievit ET AL.