Turtles all the way down? Psychometric approaches to the reduction problem
Kievit, R.A.

Citation for published version (APA):
Chapter 3

Modelling mind and matter: Reductionism and psychological measurement in cognitive neuroscience

Abstract

Can we use psychometric modelling to gain insight into the relationship between neurological measurements and psychological measurements? This rejoinder addresses various topics raised by commenters. We focus on the importance of mechanistic explanations, the nature of indicators and variables in psychometric models, philosophical theories that can be translated into psychometric models, and the notion of causality in structural equation modelling. Furthermore, we discuss extensions of psychometric models relevant for the reduction problem, and how to embed the SEM approach in recent advances in the fields of cognitive neuroscience and psychometric modelling. We conclude that our approach may bring together the best of various branches of science and philosophy that have tackled the reduction problem, and offers a fruitful and novel approach to this vexing problem.

Published as:
**Prelude**

Chapter 3 was written as a response to commentaries given by a range of scholars from various disciplines to Kievit et al., 2011 (chapter Two). Although it contains largely novel ideas and extensions of the ideas in *Mind the Gap*, it takes inspiration from the wide range of suggestions and critiques from scholars. The extracts given below summarize the core points made by the commentaries so as to make the rejoinder a self-explanatory whole.

**Commentaries**


Bagozzi positions the focus of the target article on identity theory and supervenience in the context of a larger set of philosophical alternatives, ranging from the extremes that deny the existence of the mental (eliminativism) to the idea that mental processes are non-physical and ontologically separate from the brain (substance dualism). He shows that they can all be implemented in a SEM framework that captures a broader set of theories than models in the target article.

Most importantly, Bagozzi argues that the target article neglects an important theory in both philosophy of mind and psychology, namely that of functionalism. From this perspective, what makes ‘a thought a thought’ is not its constitutive details, but what role or function it plays in a system. Bagozzi show that functionalism can be combined with a supervenience perspective, where the latent variables are partially determined by the role they play in a larger psychological whole.

Furthermore, Bagozzi argues that identity theory is most realistically captured by a two-factor model, where a non-significant deviation from a perfect correlation between the mental and the neural can be taken as provisional evidence in favour of identity theory. A deviation from a perfect correlation can be meaningfully quantified, yielding more information than mere misfit of a model-based assumption. Bagozzi argues that the implementation of identity theory in the target article is at once not strong enough and too strong, for similar reasons: method error and method bias. It is not strong enough such that by having two types of measurement, method bias could lead to inflated correlations and estimates leading to incorrect acceptance of identity theory. It is not strict enough for a similar reason: Measurement error and method bias of the implementation of either the mental or the neural could lead to misfit even if identity theory *would* hold if we were to have better measurement techniques. This is a source of concern as method bias effects are well-known in both psychology and neuroscience. These issues should be addressed not by assuming that there is no method bias, as the target article implicitly does, but by modelling it explicitly. Bagozzi further illustrates how to extend models to test convergent and discriminant validity, namely by including self-report, expert report and neural measurements of various constructs, to compare and contrast method bias (across factors), convergent validity (within factor) and discriminant (between factors). Together this explicit modelling of potential sources of method bias and measurement error can quantify important sources of error and provide more nuanced insights into the plausibility of models.
Another concern is the causal interpretation in the target article. Bagozzi argues that the causal interpretation of arrows in Structural Equation Models in general is a mistake, and that such arrows are better seen as a representation of measurement than of causality or determination. Moreover, the (causal) interpretation of supervenience in the target article is not the only possible interpretation. As Kim argues that mind-body supervenience occurs simultaneously, this negates most traditional interpretations of causality and those proposed in the target article. He notes that a simultaneous determination of the psychological by the neural can be captured in a reflective model where the correlation between the mental at time $t$ and the mental at time $t+1$ is fully explained by the neural measurements. This partially solves an issue concerning the current supervenience model as a MIMIC model: With its joint dependency on the mental and the neural, it leaves the precise interpretation of the latent variable unclear. This joint dependence is problematic for a traditional causal view on supervenience, and should be replaced by a view of the latent variable as a mathematical abstraction, not as an unambiguously causal entity. Finally, Bagozzi generalizes the approach in the target article by developing a generic model that can incorporate a wide variety of theories from philosophy of mind by setting various parameters to theory-compatible values, illustrating the broad applicability of the SEM approach to questions in philosophy of mind.


Barrett argues that the implementations of identity theory and supervenience in the target article do not exhaust the range of models compatible with these theories. Firstly, the conventional description of identity theory specifies the mental in terms of the physical or neural. As such, identity theory is not symmetrical with respect to the relative ontological status of the neural and the psychological, considering the former to be ontologically prior. In this sense, Barrett argues, it is merely our lack of measurement sophistication in uncovering the true neural state of the system that temporarily puts neural and psychological measurements on similar footing in statistical models.

There are various versions of identity theory, with the most traditional and restrictive being type identity theory. From the perspective of type identity theory, a neurological type of state (e.g. ‘c-fibers firing’ is identical to a psychological type of state (e.g. ‘being in pain’). This can be seen as a kind of biological essentialism, that assumes that there exists some biological feature $X$ that covaries uniformly with the psychological property of interest, and it is our task as researchers to uncover what $X$ is. Barrett shows that this perspective has often been, implicitly or explicitly, assumed in emotion research, with theories in the literature about emotions proposing a particular physical explanation ($X$) such that it maps one-to-one onto the emotion of interest. However, there is considerable evidence that this theoretical conception is empirically false. For instance, fear is related to at least three distinct neural systems, and none of them is the ‘real’ neural fear system, nor is fear singularly dependent on any of the individual systems (i.e. activity in each system in isolation is necessary nor sufficient for ‘being in fear’). For this reason, we need to either relax our conception of identity theory or move to a different theoretical perspective on the relationship between psychological constructs and biological realizations.
One such alternative is token identity theory. In token identity theory, individual instances (tokens) of a psychological state are identical to individual instances (tokens) of a neural state, but there need not be regularities either within or between people regarding the correspondence between neural and psychological tokens. Barrett shows that token identity theory is similar to supervenience in that there is a similar asymmetry between the mental and the physical: Identical neural states will be associated with identical psychological states, but the reverse need to be true. This conceptual shift allows us to consider the neural state of the system to be the true state of the system, where several neural states together contribute to one brain state, which in turn causes the psychological states to emerge. This emergence can take place at several levels of abstraction, from the lower neural circuits up to interacting cortical regions. Moreover, emergent properties (e.g., smaller sub-circuits of neurons) can play different roles in different supervenient properties. This many-to-one mapping has two important consequences. It solves the ontological conundrum of the how to map neural properties onto psychology: Because neural properties are ontologically 'reused' (i.e., figure in many psychological constructs), they are not exhausted by any one psychological property. Moreover, psychological states allow for disjunctive neural realizations, and it is this disjunctive property that makes the task of cognitive neuroscience so complex. Neural subsystems both stand in a one-to-many relationship to other, more complex neural states, and in a many to one relation to psychological states. This precludes simplistic inductive identification of neural identities with psychological states.

In addition to similarities between supervenience and token identity theory there is a core distinction, namely the status of the latent variable. This latent variable, or state, is realist in a supervenience conceptualization, but not in token identity. However, this is a different kind of realism than that of the individual neural states. The latent variable under supervenience is an emergent property, in the sense that it captures functional distinctions between psychological states that are useful for human perceivers in allowing for communication about mental states and human actions. Psychological categories are therefore both real and subjective, representing arbitrary divisions of the complex disjunctive causal processes underlying them. They are real in the sense that they are reliable, useful ways to carve up human experience. The ontological surplus of the latent variable above and beyond the empirical facts meaning is the cultural convention of the 'collective intentionality of a group of perceivers' that makes psychological categories useful. Most importantly, this perspective preserves the need for psychology, in capturing subjectively meaningful features of the word. This subjective realism sees psychological states as emergent properties of collectives of neural states defined by social conventions. This integrative perspective can gain insight into the complex neural underpinning whilst preserving the ontological state of psychological concepts, unifying subjective realism of the mental with the natural ontology of neural properties.


Berkman and Lieberman show that the target article implicitly assumes a dichotomy between the mental and the neural in explaining psychological
phenomena, but that this perspective overlooks an essential factor: Behaviour. The focus of both models is based around measurement and proper estimation of the latent variable on the basis of restrictive pencil and paper tests designed to measure mental processes and properties. This neglects the more direct and important task for (social) psychology: to study and predict under what situations behaviour is and is not likely to occur. Berkman and Lieberman argue that behaviour deserves to be represented as a novel, higher layer in the model hierarchy, measured by ‘B-indicators’ for at least three reasons. Firstly, the link between self-report (P-indicators) and actual behaviour is stochastic at best and biased at worst. Secondly, behaviour and behavioural tendencies vary in complex manners across contexts, and people are notably poor at incorporating situational dependency into their self-assessments. To the extent that models neglect situational variables, they will fail to accurately capture the richness of behavioural variables, and therefore lack generalizability. Moreover, for pragmatic reasons, neural processes or properties associated with real-life behaviour cannot be measured at the same time as that real life natural behaviour is occurring. This temporal gap (creating a dichotomy between ‘proximal’ and ‘distal’ indicators) between the N and P measurements will decrease the correlation between the two, biasing away from identity theoretical models. Thirdly, there is accumulating evidence that the brain does not map in a one-to-one fashion onto either the mental or behaviour even in the best of cases. Given that it is also likely that the neural processes and structure underlying behaviour differ across these contexts, these two factors interact to further complicate the measurement theoretical relationship between neural, psychological and behavioural processes. This illustrates the importance of a tripartite division by showing how neural measures can moderate the relationship between psychological processes and behaviour, and shows the conceptual and practical benefits of extending the approach of the target article to incorporate behaviour.

To summarize, Berkman and Lieberman propose to add B-indicator to the supervenience hierarchy, with behaviour supervening on psychological process, which in turn supervenes on neural processes and properties. This is in line with a wide variety of known results, such as the context-dependent nature of behaviour (suggesting various psychological processes) and the decomposition of complex behaviours (e.g. mindfulness) into a wide variety of psychological processes. Veridical modelling of the relationship between the mind, the brain and behaviour needs to incorporate all three aspects into the models, and the extent to which the former are given priority, it precludes understanding and predictive accuracy of the latter.


Burnston, Sheredos and Bechtel argue that the target article confuses data with phenomena. The former are the things we measure, the latter are the repeatable processes in the world that we aim to explain. They argue that the purpose of cognitive neuroscience is to explain phenomena of interest, not to merely model correlations between variables. By implicitly assuming the latter, the target article confuses the explanation of the phenomena with a mere (re)description of the data. By being satisfied with a well-fitting model, Burnston and colleagues argue, the target article stops at a level of description
(capturing the data) that would not satisfy practicing cognitive neuroscientists. Moreover, the identity theory in the target article is not the most suitable interpretation of identity theory. Firstly, at a low enough level of granularity identity theory will fail for both neuroscientific (people have different brains) and psychological (given enough items or fine-grained measurements, all people will differ slightly in response patterns) measurements, rendering it trivially untrue. The useful interpretation of identity theory then operates at a slightly higher level of abstraction. Moreover, it is most fruitful to assume identity theory as part of the ongoing process of this decomposition and mechanistic description of psychological phenomena. This is a different flavour of identity theory: Not an a priori truth imposed on the phenomena, but a default assumption that guides inquiry, but can be challenged on the basis of new results. This theoretical approach is called Heuristic Identity Theory. This fits better with what should be the goal of cognitive neuroscience, which is to provide a mechanistic account of psychological processes, by decomposing them into smaller cognitive processes each associated with individual neural processes. Burnston and colleagues illustrate the effectiveness of HIT in analysing processes involved in face perception, where assuming a one-to-one mapping of neural activity related to face perception has led to increasingly detailed insights, whilst simultaneously continuously revising the original, simplistic identity theoretical relationship between, for example, the Fusiform Face Area (FFA) and face perception. This approach of varying experimental input generates identity theoretical hypotheses, which can then be tested by manipulating the processes of interest, for instance by interrupting neural processes by means of TMS.

Finally, they propose a reductionist framework that focuses on decomposing psychological processes into smaller units that can be described as forming parts of the process being understood. This approach allows for the simultaneous study of phenomena at different levels, and can capture the seeming paradoxes of top-down versus bottom-up causation. Processes that look like top-down causation by studying the details of the (simultaneous) processes occurring at different levels. Ultimately then, the gap proposed in the target article will disappear: processes and phenomena at different levels of description can be translated to other levels in such a way that they can be understood as descriptions of the same phenomena.


Haig positions the approach of the target article in a broader philosophical context, and shows that considerations from philosophy of science, in addition to philosophy of mind, are pertinent to approaching the reduction problem. Firstly, he shows that the underlying philosophical background of the target article is best seen as a strong version of naturalistic realism. Naturalism is the position that considers science and philosophy to be ends of a continuous spectrum of theoretically guided empirical investigation of the world. In this sense it is a thoroughly realist view of science in line with the epistemological naturalism of Quine, who argues that philosophy is in many ways similar to, and continuous with, science. This perspective suggests that philosophy as a hypothetico-deductive method of inquiry into the nature of reality is not inherently different from science, and that empirical science does not need to
answer to any a priori philosophical considerations— they represent the ends of a continuum. However, according to Haig, beyond Quine the target article argues that in addition to philosophy providing theories and conceptual foundations that can guide empirical inquiry, results from science can and should constrain philosophical theories. By approaching the reduction problem of the mind and brain as a question of model testing and model fit, this point of mutual constraint is made explicit: Theories from philosophy of mind can be rejected or constrained if the predictions are not supported by the data, and cognitive neuroscience should be aware that there are no philosophically neutral approaches to empirical attempts to link mind and brain. The target article implements this approach by means of Structural Equation Modelling, which is in essence a hypothetico-deductive method. That is, the predicted covariance matrix on the basis of the (philosophical) theory is compared to that observed empirically, and the deviation from this predicted covariance, quantified by means of fit indices, is taken to be the arbiter in judging the success of the theory. However, Haig notes that this neglects other approaches to empirical and explanatory adequacy. One such approach is inference to the best explanation, from which perspective the degree of explanatory coherence is central, which states that statements and propositions that a part of a theory should tie together closely and generalize beyond the current, contingent, data. This can be implemented, for instance, by focusing on measures of model parsimony and comparative model fit that weighs the complexity of the model in relation to its empirical adequacy. A justification is that overly complex model will do poorly in terms of generalizability and explanatory coherence, as they will be too heavily weighted towards explaining locally contingent phenomena, and thus are not good candidates when considered from the perspective of inference to the best explanation. Together, the synthesis of naturalistic realism and inference to the best explanation can serve as a philosophically guided approach to the reduction problem.


Vul (2011) takes an explicitly pragmatic approach to the reduction problem, based on three arguments. Firstly, there is no a priori principled reason to consider any level of abstraction to be primary. Secondly, models at all levels of explanation are likely to be wrong. Thirdly, explanatory models always carry specific costs in terms of time and effort and benefits in terms of the accuracy and granularity at certain explanatory levels. We should then ask of reductionist theories: What is gained, in terms of prediction, accuracy and insight, by studying a higher order phenomenon on the basis of lower order realizations? It is the answer to this question that should guide us in deciding at which level in the hierarchy we should ‘stop’. Based on this principle, we have at least some justification for cognitive neuroscience, as it has led to certain useful level-crossing predictions (e.g. guiding neurosurgery). However, from this perspective, reductionist models have an inherent challenge: Models and techniques developed at particular levels of explanation (be they psychological, neurological, molecular or other) are generally more useful and more cost-effective at predicting and explaining phenomena at that level of explanation. The only principled way of arriving at a unified reductionist psychology, Vul argues, is by developing computation models of cognition and the brain that simultaneously predict phenomena at both levels of abstraction.
Rejoinder

According to Karlin (1983), “The purpose of models is not to fit the data but to sharpen the questions”. 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, empirically tractable, and can be improved and revised on the basis of empirical and conceptual advances. Most importantly, 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 formulated by Kievit and colleagues (2011), based on the ideas and suggestions offered by the commenters. Our focus will be on the following set of themes that figured centrally in the comments: 1) What is the role of mechanisms with respect to our approach; 2) What explanatory levels should we study; 3) Why we should engage in reductive science in the first place; 4) How can psychometric models be extended; 5) What interpretations of causality are relevant for psychometric models and 6) Which other philosophical positions can be translated into measurement models?

Mechanisms and measurement models

Perhaps most critical of our enterprise are Burnston, Sheredos and Bechtel (2011). They summarize their criticism as follows: “Our central objection to the psychometric approach deployed by Kievit and colleagues 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” (italics in original). 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 (behavioural genetics, molecular chemistry). However, despite our concurrence with this overarching goal, we doubt whether the arguments of Burnston and colleagues (2011) are relevant to our central point. First, the two endeavours - discovering mechanistic explanations and formalizing correct measurement models - are not mutually exclusive goals in science. In fact, they are mutually reinforcing. Hence, that mechanistic explanations are important does not imply that measurement models are unimportant or vice versa. Second, the quality of mechanistic interpretations is essentially dependent on the quality of the measurement theoretical foundation used to relate psychological and neural phenomena. This is simply because the quality of any empirical study depends on the quality of its measurements.

We consider the second point first. As Bagozzi puts it: “Substantive researchers do not always develop the conceptual meaning underlying measurement relations, but they are implicit in the theoretical development of hypotheses and deserve explicit consideration.” 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 on correctly specified measurement models also holds within psychology: when we find a consistent, replicable difference in performance across participants in a
particular test (e.g. perceptual organization), we would ideally like to explain why this is the case. However, this question is secondary to, and reliant upon, the question of what the phenomenon (perceptual organization) is, how we may measure it reliably and validly, and whether our measurements allow for correct insights that form the basis of mechanistic explanations.

This dependency is especially relevant for neuroscience. For instance, consider the research discussed by Burnston and colleagues (2011). 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 papers (that it may be better considered an area selectively active for stimuli for which a person can be considered an expert, for instance 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 (BOLD)-response (Logothetis, Pauls, Augath, Trinath, & Augelman, 2001), inter- and intra-individual variability in the Hemodynamic Response Function (HRF, the form of the temporal response of the increase in oxygenated blood) (e.g. Handwerker, Ollinger, & D’Esposito, 2004), 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 above shows, these issues are well-known and well-documented in the field of cognitive neuroscience, and practicing researchers (such as Kanwisher et al.) are generally well aware of their importance. For instance, if we were to (incorrectly) assume that the HRF is uniform across and within individuals, it may be that we discover spurious (de)-activation in certain regions that is actually an artefact 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. it 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 behavioural 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 and colleagues rightly praise the advances and refinements of the FFA after the initial findings by Kanwisher and colleagues (2003). However, the refinements of the original FFA theory based upon subsequent research (indeed an excellent example of Heuristic Identity Theory) can also be tackled by means of correctly specified psychometric models. For instance, research based on the original study by Kanwisher and colleagues showed that, contrary to the initial hypothesis, the FFA also shows differential activity to non-face stimuli (e.g. to other visual stimuli in which a person can be an expert, such as bird perception). In Figure 3.1, we show that this fact could have been detected by means of a correctly specified psychometric model.

---

**Figure 3.1.** Using structural equation modelling to examine the presumed function of the Fusiform Face Area (FFA). Note. Here we have two latent traits: whether an individual perceives a face (measured by P- and N-indicators), and whether an individual perceives a bird (assuming expertise in bird perception, after the examples discussed in Burnston, Sheredos, & Bechtel, 2011). If activity in the FFA is truly only selective for face-stimuli, then activation in the FFA should be determined by the latent trait “bird perception.” If we specified this model, then a significant crossloading between the FFA and both latent traits would lead us to revise the original hypothesis concerning the FFA.
The refinement of the theory could have been detected by showing there exist significant cross loadings of the FFA to two distinct latent variables (e.g. face perception and bird perception). That is, if the activity in the FFA is not simply a measurement of face perception, we expect the cross-loading (\(\lambda_{12}\)) 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 cross loadings 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 and colleagues 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 is 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 and colleagues (2011): 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? As Vul (2011) and Barrett (2011) point out, the neurological measures we analysed in our empirical illustrations (Kievit et al., 2011) represent an arbitrary, and perhaps suboptimal, point to ‘stop’, and argue that those measures are too coarse to be plausible candidates for reduction. Interestingly, arguing that the level we

---

10 Although both Burnston et al. (2011) and Vul (2011) 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 note that our approach is compatible with the algorithmic or computational level, more on the flexibility of N-indicators in a later section).!

11 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. 138): Despite the
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 in the first place?

Vul (2011) discusses a possible way to judge the success of reductive science in connecting different explanatory levels: 1) a main goal in cognitive neuroscience is prediction (e.g. the predictive success of the accuracy of a surgery induced lesion), and 2) given this criterion of prediction, we judge the success of a specific indicator (e.g. grey 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, i.e. when brain measurements do not outperform behavioural measurements of a given psychological construct.

For instance, Burnston, Sheredos and Bechtel (2011) 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 behavioural 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 (2011) takes a broader perspective concerning 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 (2011), Vul (2011) and Berkman and Lieberman (2011) 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 (2011) and Berkman and Lieberman (2011), who argue coarse granularity, these measures together explain 25.1% of the variance in g, which is the higher end of explained variance in most cognitive neuroscientific research papers.
that modelling of reductive science can be used to quantify and possibly overcome method bias in a new way. This is an excellent suggestion that we will return to later.

In addition to this discussion on the specific utility of cognitive neuroscience, Haig (2011) 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 super-empirical virtues in science – super-empirical 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 endeavour. For instance, if the implicit value by which we judge scientific explanations is parsimony, then we may judge if, by this criterion, a neuroscientific description of a psychological trait succeeds. 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 is adequate, no matter how much it lacks in elegance or simplicity. Similarly, if the super-empirical value is the unifying power of a theory, then an overarching theory may lack in 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 behavioural ecology, is as large as one could wish from a scientific theory (of course the above 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 super-empirical values by which one wishes to judge science in general, as well as the perspective from which one engages in reductive (cognitive neuro-) science determines the framework from which one judges the adequacy of a particular line of research. There is no single correct perspective, and this fact alone implies that the assumptions of reductive (neuro)science should be made explicit in research to ensure that its relative success is judged on the basis of the appropriate merits.

**Brains, people and reductionism**

It is not uncommon in both the empirical and philosophical literature to find the following 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.” (bold emphasis added). Strictly physically speaking, this is of course true. However, in terms of explanation, this need not necessarily be the case. Kim make a similar point concerning physical reduction: “The core of contemporary physicalism is the idea that all

---

12 Haig (2011) 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).
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, cited by Bagozzi, 2011). 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 will 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, e.g. 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 argue 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 makes a similar point in 2008, 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. 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 eliminative materialists such as Churchland, as discussed by Bagozzi, tend to argue that psychological phenomena are not ‘real’ but are ‘actually’ neural or hormonal properties such as hormonal or neural 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 pragmatic reductionism account. This account offers us grounds to stop at the explanatory level of the brain (e.g. grey 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, p. 67): “Our mutual understanding and even our introspection may then be reconstituted within the conceptual framework of completed neuroscience” (emphasis added). 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 favour 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.’” (pp. 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: 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 offered 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 (2011), Burnston, Sheredos and Lieberman (2011) and Barrett (2011) argue, based on the indicators we used in our analysis, that it is unlikely that activity in a single brain region, or grey matter density in a particular area, will ever be identical to a complex trait such as intelligence, and are 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 modelling in this way.

*Figure 3.2.* Supervenience “deluxe.” An extended supervenience model. P-indicators are only drawn from the first psychological variable for visual clarity. This model represents hierarchical multiple realizability. The same behavioural state (e.g., “freezing behaviour") can be determined by different psychological latent trait values, which in turn can be determined by different constellations of neural activity and/or constitution.
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 wish to remedy that problem 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 grey 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 measurement 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 and colleagues (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, i.e., 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 are 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. 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 (2011) argue that instead of explaining phenomena, we are only explaining data. 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 not explaining 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 (2011), “all measurement questions are also philosophical questions about how variation in numbers hint at or point to reality”). 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 towards 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.

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, 2011, 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 (for instance, 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 co-varying neurological measurements, which if it correlates 1 with a psychological variable, can be considered as evidence for identity theory. 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
indicators ‘by themselves’ by means of an exploratory factor analysis may do more justice to our current prevailing ignorance concerning the structure of neurological covariation. This is an excellent suggestion and an approach we have followed and implemented in more recent work (Kievit et al., 2012), where we examine intelligence and a more complex constellation of brain measurements than those modelled 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 latent variables of interest. Using an exploratory factor analytic approach for different constellations of neural indicators may provide novel insights into the structure of inter-individual differences in neural makeup.

Barrett discusses a wide range of new models, largely from the perspective of psychological construction (Coan, 2010; Barrett, 2009, 2011), 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 and psychological components of emotions. 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 whilst 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 behaviour**

Berkman and Lieberman (2011) argue we should include measurements of behaviour in our models, significantly extended the scope of our approach. They argue that behavioural 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 behaviour and psychological characteristics. This importance of behaviour in such models, and the role of multiple determination is also suggested by Barrett and Bagozzi. Barrett, for instance, discusses several examples where the one type of overt behaviour can be caused by different constellations of (presumed) psychological states. For instance, in animal behaviour research, certain behaviours (e.g. freezing behaviour) 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 (2011), presented in Figure 3.2, could be implemented to examine the predictions concerning determination and multiple realizability of behaviour with respect to psychological indicators and neurological indicators in the same model. The constellation of psychological components that lead to a given behaviour is generally quite complex. If only the behavioural outcomes are studied, we may not be aware that different interventions that are different from having two separate, albeit highly correlated, latent factors.
constellations of psychological primitives can lead to the same type of behaviour. Similarly, by only studying psychological processes or dispositions, we may be unaware of the constructive role they play in behaviour, thus this yields benefits for the two fields of study. This interpretation may be referred to as a hierarchical supervenience model, and is similar to the constructionist model Barrett (2011) proposes in Figure 7B.

Models like the above 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 the interactions in 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 genetics loci, Penke, Dennissen and 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, behaviour 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 fruitful empirical approach that can serve to link different manners of inquiry.

Mismatch of latent variables: Sources and solutions

Berkman and Lieberman (2011) argue for the inclusion of behaviour 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 behaviour) do not match observable or measurable behaviour (objectively observed ratings of behaviour). The issue of measurements of the same construct not being perfectly correlated is a more general problem that also applies when comparing neural and psychological levels (consider, for instance, Bagozzi’s proposal to model a latent variable with only neural indicators and a latent variable with only psychological indicators, and the correlation deviating from 1). Given our focus on integrating different explanatory levels, such mismatches are worthy of close attention.

There are a variety of reasons for a mismatch between latent factors. We will discuss four reasons that are especially relevant for our approach and can be addressed using psychometric models. Firstly, 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 behavioural 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 noting the behaviour 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’ behaviour 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 also a factor score based on a measurement tool that measures behaviour in a set of relevant social situations (‘Extraversion behavioural report’, 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-behaviour’, both of which may be psychologically relevant (e.g. the first with respect to theories of self-perception, the second for dominance theories, or theories of group behaviour). 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. 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 behaviourally 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 2011) suggest. The influence of the environment can be modelled 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 1 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 behaviourally observed extraversion in a certain setting. The influence of such environmental variables can easily be incorporated in a psychometric model, as is illustrated in Figure 3.3.
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 more shy in a high-pressure environment) or a host of other psychologically and behaviourally relevant possibilities. We can test this hypothesis with a multi-group 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 behaviour) that influences the latent behavioural trait. In this manner it is possible to quantify the influence of the environment on the behavioural factor, and to partially explain the fact that two latent traits do not correlate perfectly. Not only does this method increase insight, it also allows us to formalize explicit hypotheses concerning the status of psychological constructs and how we expect them to (co)vary, and refine our insight into the sources of mismatch between psychological and behavioural properties. This again represents a case where psychometric models not only allow us to model certain plausible disturbances, but also makes the measurement theoretical assumptions of the researcher explicit.

A final explanation for the mismatches between latent variables considered here 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. 2006) has argued that psychological measurement of consciousness or conscious awareness is...
inherently problematic due to unavoidable bias in any psychological measurement. In consciousness research, almost every experimental paradigm takes self-report as the gold standard. That is to say, in such experiments, specific parameters of stimuli, usually visual (e.g. clarity, presentation time) are manipulated. Subjects are asked whether or not they perceive the stimulus, and when subjects report ‘yes’, such trials are compared in terms of neurological measurements to trials where subjects 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 or not 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, since it takes a couple of hundred of milliseconds to (verbally) report awareness of 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 range of visual experience.

**Figure 3.4.** Estimating method bias inherent in self-report. Here we can test the hypothesis that the error terms of the psychological assessment of the latent trait are correlated. We can free the parameter and use a likelihood ratio test to test whether the fit of the model improves significantly.
We propose a model to test whether self-report is biased with respect to consciousness of a stimulus in Figure 3.4. The purpose is to quantify whether biased self-report leads to correlated error terms, and so to poor measurement of the latent trait. In Figure 3.4, the psychological indicators represent self-report or behavioural 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 variable ‘bias’, an exogenous influence that directly affects the correlation between psychological indicators of awareness.

This variable, representing measurement bias, is the proposed cause of correlated residuals between psychological indicators, for instance because people forget what they ‘actually’ perceived. 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 non-random 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 3.4. 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). In this framework, 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 they may not suffer from bias.

Realism and naturalized philosophy
Several commenters have 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, Sheredos, & Bechtel, footnote 4). It would be unwisely ambitious to attempt to resolve the ontology of psychological attributes here in one fell swoop, and we will not attempt this within the confines of this paper. 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 temperature. 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
Psychometric approaches to the reduction problem

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 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 (for instance, in the case of intelligence, it is possible that one concluded that John is more intelligent than Jane, while 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: Its must be able to function as a truth maker for hypotheses that bear on it. As MacCorquodale and Meehl (1948, p. 106) say, such attributes "have a cognitive, factual reference in addition to the empirical data which constitute their support". Is such a factual reference sufficient to call an attribute ‘real’? That is a question on which debate is possible. Consider Dennett (1991, pp. 27) on the question of whether ‘centres 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 semi-realism. …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 meta-physical and the scientific.”

Centres of gravity would seem to have pretty good credentials compared to most psychological constructs proposed today. They are measurable, determinate, one can be wrong about centres 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 we can justify the assumption that psychological attributes are determinate in this sense? Can they function as common causes? These are open questions. We cannot resolve them; but we can make an interesting observation on why they are open. Namely, in many cases of psychological testing we do not have a clear idea of the processes that lead to test scores, and hence it is hard to trace back these processes to the level of the attributes measured. The introduction of psychological attributes as latent variables in reflective measurement models thus rests on something of a gambit; namely that someday someone will be able to flesh out the processes that connect the observables to the attribute of interest. How to flesh out such process level explanations is an active area of research (e.g., see Van der Maas, Molenaar, Maris, Kievit, & Borsboom, 2011). 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 does not, in itself, 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 is a productive assumption, and most compatible with current scientific practice. However, others do not necessarily have to agree with our perspective to be able to benefit from the possibilities offered by the SEM-framework. There is a wide range of perspectives on the reality of (latent) constructs both within the psychometric community and the psychological community, and researchers from different perspectives may still employ the same techniques, despite having diverging assumptions concerning the reality of constructs.

**Causality and structural equation models**

Several commenters discuss the precise nature of the causal relationships we assume in our models. The central issue of this paper, 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). For a discussion on different statistical interpretations of mechanisms and causal intervention, see for example Holland (1986) and Rubin (1986). 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). 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). As is
Psychometric approaches to the reduction problem

clear from the above, a wide range of causal perspectives are possible and can figure as implicit foundations of empirical inquiry. However, we do not wish to not 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 (Horgan, 1993; Kim, 1987; Lewis, 1994). Burnston, Sheredos and Bechtel (2011) for instance argue that supervenience is a relationship of realization, not of causation, whereas Barrett (2011) describes a causal interpretation of supervenience, drawing on Searle (1992). Even from the same author, Kim, such a wide range of interpretations can be drawn. For this reason, as we noted 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 and colleagues 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 one 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 of 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 $\zeta$ on the latent variable represents the totality of unmeasured causes not included in the current, incomplete version of the
model. Diamantopoulos (2006, p. 14) 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.” 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 from the above that there are many sensible positions concerning the causal relationships in the various models, and we do not presume to favour 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 discussed in Kievit and colleagues (2011), inter- and intra-individual explanations differ quite radically. To summarize: In our view, the use of models such as the reflective and formative models discussed above 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 original paper, 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 will 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 will examine in more detail. Identity theory, as supervenience, comes in many flavours, compatible with different measurement models. The most important distinction is between type and token identity (e.g. Aydede, 2000; Fodor, 1974; Place, 1999). These two positions 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-fibres firing’ (Y), or X iff Y. That is, if this relationship can be said to hold, for any individual that is in pain their C-fibres must be firing, and when C-fibres 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).
Psychometric approaches to the reduction problem

Token theory 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 of consistency on this relationship: in the same person, the same psychological state X may be identical to a different neurological state Y2 at a different point in time. This position then, as Barrett notes, carries no real ontological weight, and is better seen as an anti-dualist grounding of psychological states in physical reality. We then have two versions of identity theory, 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

Token-token identity theory, as Barrett notes, is consistent with some interpretations of (local) supervenience. Consider Kim (1998): “For example, if a person experiences pain, it must be the case that that person instantiates some physical property (presumably, a complex neural property) such that whenever anyone instantiates this physical property, she must experience pain” (p. 9). However, both type-type and type-token identity theory are, in our view, not compatible with supervenience.1

---

14 Token-token identity theory, as Barrett notes, is consistent with some interpretations of (local) supervenience. Consider Kim (1998): “For example, if a person experiences pain, it must be the case that that person instantiates some physical property (presumably, a complex neural property) such that whenever anyone instantiates this physical property, she must experience pain” (p. 9). However, both type-type and type-token identity theory are, in our view, not compatible with supervenience.1
some identity theoretical relationship at any given time. We may then consider an alternative in between these two options, which we will 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 this model, for me, is not the same as for someone else, this means the group-level type identity model will not fit.

Measuring fluctuations in of a suitable constellation of (P and N) indicators over time may 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 constellations of A’s, B’s and C’s.

There are substantive and philosophical reasons to consider this interpretation of type-token identity as a viable candidate. Firstly, as both Vul (2011) and Barrett (2011) note, 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 neural level at all for which, for instance, a reflective model can be said to hold. Here we part with Bechtel and Mundale (1999) who state 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) for a given psychological construct exists 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 identity can be said to hold, then we would consider the type identity description more parsimonious and therefore preferable. If a 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 2.5, 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 (‘degree of’) 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. Figure 2.5 shows how we can specify an identity theoretical model of a psychological construct for both individuals, with different parameter estimates. Inter-subject variability can then explicitly be modelled, and individual differences in neurological measurement models may be described.
This would be in line with Frege’s remark “He does not have my pain and I do not have his sympathy” (Frege, 1956, p. 300): Although it may be possible to construct an identity theoretical model concerning, say, pain or depression and its neurological realizations for every individual, it need not be identical for every individual, but may still be able to provide valuable insights. This same structure may also hold for Berkman and Lieberman’s (2011) treatment of the multiple realizability of psychological properties with respect to behaviour: for each individual, an identity theoretical relationship between psychological properties and behaviour 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). These 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 Hale 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 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. But it may still be the case that all people, at least those with sufficient cultural exposure to the above 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 to 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 the above 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 paper by Kanwisher, McDermott, and Chun, (1997), the data analysis yielded different (but similar) neuroscientific findings for the 15 subjects 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 less parameters) model, and as such is preferable.

**Emergence and mutualism**

A final theory we will consider for its empirical and conceptual implications is emergence. It is briefly mentioned by Bagozzi (2011), and in more detail by Barrett (2011). A general description of emergence is “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, pp. 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 cannot, it seems, be fully 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 by Van der Maas and colleagues (2006), who showed that 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 grey 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 distinguish between hypotheses that, with cross-sectional data, cannot be differentiated (cf. Barrett, 2011). That such time series modelling of brain measurements through time may yield counterintuitive results was illustrated by Shaw and colleagues (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 positive for the same group at a later developmental stage. It is these types of developmental phenomena that fit 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 GLM designs, despite the fact that some of this undetected neural activity may be of importance for the task of interest.

Future developments
So how should we move forward? Firstly, we think the models proposed by the commenters and 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.

Secondly, 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 modelling. 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 datasets such as we discussed in the type-token model example Figure 2.5 (Hamaker, Nesselroade, & Molenaar, 2007), and Bayesian implementations of latent variable models possible (e.g. Congdon, 2003, chapter 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, 2011). This project aims to structure the knowledge concerning psychological constructs and 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 where 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 grow from an ‘adolescent field’ (Kriegeskorte, 2010) into an adult field, it needs to formalize its models has been steadily gaining traction (e.g. Henson, 2005; Just, Cherkassky, Aryal, & Mitchell, 2010; Mars, Shea, Kolling, & Rushworth, 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, 2011), we have hopefully shown the benefits of treating it as such. We
have argued above that the extent to which the brain is informative with respect to specific psychological properties should be considered an open question. For the time being, we will take Barrett’s (2011) advice: “For the present, however, it makes sense to forge ahead….”. 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.