

## Reductionism and Practicality

Edward Vul

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

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

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

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

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

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

In the target article, Kievit et al. (this issue) describe a psychometric approach to connecting cognitive

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

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

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

### Note

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

### References

- Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in statistics* (p. 202). New York, NY: Academic.
- Marr, D. (1982). *Vision*. Cambridge, MA: MIT Press.