THE "RUTHLESS REDUCTIONIST"

A CONVERSATION WITH JOHN BICKLE

by Anders Strand

John Bickle is Professor and Head of the Department of Philosophy and Professor in the Neuroscience Graduate Program at the University of Cincinnati. He is the author of *Psychoneural Reduction: The New Wave* (MIT Press 1998), *Philosophy and Neuroscience: A Ruthlessly Reductive Account* (Kluwer academic Publishers 2003), *Understanding Scientific Reasoning*, 5th Ed. (co-authored with Ronald Giere and Robert Mauldin, Thomson-Wadsworth 2006), and numerous articles. He is also the founding editor of the *Studies in Brain and Mind Book Series* from Springer Publishers.

Bickle is best known for his work on, and defense of, scientific reductionism, especially as it pertains to contemporary neuroscience. Bickle visited Oslo during the conference *Neuroethics and Empirical Moral Psychology* at the University of Oslo in March. I started out by asking him about his impression of the philosophical attitude towards recent developments in neuroscience.

Do you think philosophers are sensitive enough to advances in contemporary neuroscience?

When I was in graduate school, 20 years ago, philosophers' interest in neuroscience was certainly much smaller than now. I think there has been a significant growth, not only among younger philosophers; older philosophers, who have been in the profession for a long time, have also gotten onboard. However, one of my main worries is not whether philosophers are paying sufficient attention, but are they paying sufficient attention to the broad range of neuroscientific results and studies. Unfortunately, the answer to that question is that they haven't. Even philosophers who are sensitive to neuroscience, and who use neuroscience in their arguments, still mainly concentrate on cognitive neuroscience. I'm convinced that cognitive neuroscience, as a discipline, is less reductionistic than some of the other branches of neuroscience. It is really interesting that some of the few philosophers and cognitive scientists who have paid attention to cellular and molecular neuroscience, like Ken Schaffner and myself, are also the ones who have advocated reductionistic programs in philosophy in general.[1]

You're taking on the task of informing philosophers of cellular and molecular neuroscience then?

Exactly. Educational outreach, as I often say; mainly as a joke but there is something really serious behind it. There is work in these other branches of neuroscience, branches much more prominent within neuroscience itself than cognitive neuroscience, that has philosophical implications and ought to be better known among philosophers. We shouldn't just single out philosophers either; cognitive scientists don't know about this work for the most part.

So there are knowledge and interest barriers even within neuroscience itself?

Yes, but I think it is natural that philosophers and cognitive scientists would concentrate more on cognitive neuroscience. The issues being addressed in cognitive neuroscience are closer to the kinds of concerns that philosophers have about the mind. The kinds of cognitive phenomena that cognitive neuroscientists are beginning to explore, as we've seen here at this conference, are certainly much closer to the concerns that philosophers of mind have focused on than work on memory consolidation in a rodent model. It is also the case that the methods of computational neuroscience are closer to the methods that you learn about in your philosophical education. The methods of cognitive neuroscience, as well as the concerns, are just closer to the kind of general education that philosophers have than the methods of molecular biology, molecular genetics and cellular physiology that dominate the other areas of neuroscience. A philosopher or cognitive scientist who's going to take on that literature is going to have to reeducate him or herself in an area of science that he or she most likely hasn't been exposed to. So there is this barrier, but my claim is that the payoff of trying to bridge it is worth the effort.

One of the main lessons you draw from your interest in neuroscience is reductionism. I would like to focus on some aspects of the reductionist program you propose. In his 1998 book Mind in a Physical

World, Jaegwon Kim started off by noting that being labeled a 'reductionist' is anything but a positive description…[2]

Yes, I've even cited that quote before.

The question is, then, do you think that this description still applies to the general philosophical attitude towards reductionism?

I go around and give a lot of talks, and a lot of times I'm asked to go to conferences in order to "represent the reductionists". I think people like to have somebody offering that perspective as a participant in the debate. Nevertheless, I still run across people who say "You're the only reductionist we know" and "Wow, it's strange to meet someone who's really a reductionist", so I think Jaegwon put his finger on something that is certainly correct within philosophy. Reductionism still has some negative connotations, probably fewer than it did 25-30 years ago, and there are still very few people who will say "I'm a reductionist" or "Here's my account of reduction". It is certainly a minority view within philosophy of mind and even within cognitive science generally. However, as we saw when Johan commented on my talk at the conference, most practicing neuroscientists who work at the level of cellular and molecular mechanisms just take the kind of reductionism that I was offering for granted; it is part of their practice.[3] It is interesting to see this disconnect between the attitudes towards reductionistic work within different branches of inquiry.

Sometimes I get the impression that reductionism has become less unfashionable in just the last 3 or 4 years.

Yes, I think that's right. One of the things I've always tried to stress is that reductionism in actual scientific practice is different from what philosophers have always thought reductionism was. The last few years I've tried to emphasize that even more. I hope that the new accounts of reduction, that are not sort of straightjacketed to the old Nagel model, will enable philosophers of mind to reconsider these newer pictures of what reductionism is in the context of their concerns.[4] Even just a decade after Nagel's book there were few people in philosophy of science who advocated Nagel's model as a correct account of reduction in science. However, if you go back and read the debates in philosophy of mind in the mid- nineties, it was just assumed that reduction was to be understood on the Nagel model. Everybody would give lip-service to weaker accounts, but would then just turn around and claim that multiple realizability shows that you cannot have bi-conditional bridge laws.[5] Actually, as my colleague Bob Richardson pointed out 25 years ago, Nagel's model itself didn't even require bi-conditional bridge laws.[6] So there's been a widespread misunderstanding of what reductionism is that has been prevalent in philosophy of mind for too long. Hopefully, loosening up the notion will make people realize that reductionism actually has something to contribute to these debates. It is important to realize that we, the new reductionists, are not talking about the same notion that was criticized 30 years ago.

Can you give a brief outline of your account of scientific reductionism?

Sure. Reductionism in scientific practice is a matter of intervening causally into increasingly lower levels of biological organization. One can for example intervene causally into particular neurons. Then one uses knowledge of the neuroanatomical circuitry to get from the particular neurons you intervene into and out to the motor periphery to track the effects of these interventions. Then you might get statistically significant differences under controlled experimental situations. This approach constitutes reductionistic practice in cellular and molecular neuroscience. Ideally, you both intervene to decrease the hypothesized mechanism and get decreases in your behavioral measures for the cognitive phenomena you study, *and* you intervene to increase the hypothesized mechanism and get increases. In practice, the second kind is much harder to achieve in mammals, but there have been a variety of successful interventions of that sort.

It's important to emphasize causality here. This is not merely a matter of correlating occurrences of particular molecular events in neurons with behavior; this is intervening causally to change those hypothesized mechanisms and eliciting changes in behavior. So it's quite different from the kinds of measures that dominate contemporary cognitive neuroscience, which most cognitive scientists admit are merely correlations.

Most of this research is done on mice. A worry is whether the results transfer to humans. Moreover, why think that results concerning quite limited cognitive functions underwrite a reductive understanding of more complex cognitive phenomena?

That's a perfectly good question, and one that I often anticipate in my talks. It's not just a question about "Ought we to be experimenting on rodents in order to learn something about the human mind?" It's also a question that has to do with one of the traditional concerns in the philosophy of mind; namely *multiple realizability*. Maybe rodent cognition has to do with a particular set of molecules, but why think that human cognition has to do with the same set of molecules? One of the benefits of going to the cellular and the molecular level of research is that neurons and intra-neuronal processes in mammal brains have not changed very much. And there is a very good reason for that. Most of those cellular and molecular processes have to do with either neuronal activity: action potentials, field potentials and the like, or they have to do with basic cellular "housekeeping" activities. It is just a matter of fact about the particular proteins and genes that perform those functions that there is very little room for any kind of changes in their functional domains. For example, if you manipulate a particular amino acid on those proteins you can create a three dimensional configurational difference in the way that the protein folds, and this difference can make it impossible for it to do its functional job within the cell. There's very little room here for evolution to work. That's why, when you get to the level of cellular and molecular neuroscience, a mouse neuron is a primate neuron is a human neuron, with some differences of course. For example, the CREB molecule, at least in the functional regions of that protein, is literally amino acid for amino acid identical in primates and in humans. And it gets even more surprising than that. Go back much further in evolutionary time. Go back to the ancestors that we mammals share with the gastropods and the insects and you'll find literally amino acid for amino acid sequence identity in those functional regions of the protein. When it comes to those particular mechanisms, at least in real world creatures, there simply is not very much multiple realizability.

A harder question is "What about the behavioral measures that you're using. Aren't you too limited in a rodent model to capture the full range of human memory consolidation?" The quick answer is "yes, you are limited". However, human beings can be tone-shock associated, and human beings can be contextually conditioned. If you put a human being in an opaque pool of water that was too deep for them to stand up in and the sides were too slick for them to grab onto, they'd search for a platform pretty much like a rodent does. If you had appropriate visual stimuli on the wall they would use those visual stimuli to find that platform the next time you dumped them in this imagined human version of the Morris water maze.[7] These particular tests, while admittedly limited to the kinds of tests we can do in a rodent model, are also tests of phenomena that are real human phenomena as well. Couple that

with the evolutionary conservation of these molecular mechanisms, and I think you can make the case for how we can generalize from what we learn in rodents to what we hypothesize in humans.

For several philosophical concerns we individuate actions more finely than as bodily movements. For example, differences in intentions might reasonably be said to give rise to differences in human actions, while the bodily movements might very well be identical. Why do you find it unproblematic to individuate actions as bodily movements?

For the purposes of this kind of research you usually individuate behavior in terms of muscle movements against the calcium frame that moves the skeleton and the movement of the skeleton through space. On this account, one particular set of bodily movements will count as the same action as another even if it's in a different context. This is not only a terminological dispute about the term "action". The way we use this word within larger research programs, and the way we evaluate who has got the best account of the individuation of behavior, is by seeing who has the most explanatorily powerful framework that that account is embedded within.

Then one might ask about these broader philosophical concerns, such as where basically the same movement of the body through space on one instance can act as signaling a score in a game and in another context constitutes signaling a taxi cab. Are those the same behaviors? Admittedly, empirical evidence starts to flag at this point, because we haven't built enough neuroscience to show anything conclusively. However, I'm convinced that once you start tracing the causal mechanisms from the production of that behavior back into central areas you will find a causal difference in the processes that generated that same behavioral movement in the different contexts. That is, I don't think we have to get all the way out into the environment to find this difference, I think we'll find it in the brain.

But it seems to me that a worry remains as long as we individuate the explanandum, the action, finely enough. Consider an example from the philosopher's armchair; Putnam's Twin-Earth example. In this case we rule out brain differences at the outset. Oscar lives on Earth, and his twin Twin-Oscar lives on Twin-Earth. Earth and Twin-Earth are identical except for the fact that water isn't H_2O *but XYZ on Twin-Earth. XYZ is functionally similar to* H_2O *but structurally different. Imagine that Twin-Oscar is visiting Oscar on Earth, they go to a café, and they both ask for a glass of 'water'. Is that the same action? As a result they both get a glass of water* (H_2O)*, but only Oscar got what he asked for. In this case it is, by assumption, no internal difference between Oscar and Twin-Oscar. The point is that, if you think there are reasons to distinguish these effects in terms of these kinds of intentional differences, then it seems like you have to bring in external factors in order to explain how those differences were brought about.*

That's an excellent example, and one that I think is insufficiently appreciated by practicing neuroscientists. I mean, you start talking about these Twin-Earth cases and their eyes just glaze over. However, in the example you've not only got mental states sort of extending out into the environment; you've also got behavior and behavioral effects characterized in terms of semantic properties. The question is what can you do with that account of the mental independently of answering contrived philosophical questions? What does it gain you explanatorily? What kinds of cognitive phenomena are now going to be explainable with this broader conception of behavior and of mental states that are not explainable on my more restricted picture of behavior and the appeal to internal causes of that behavior?

It could, for example, matter for certain moral concerns whether an action is partly caused by a true belief rather than a false belief, or by a well justified belief rather than a poorly justified one.

I can think of plenty of scenarios where my approach will be able to give an explanation that the one you're proposing will simply have to piggyback on and accept. Moreover, I'm having difficulty thinking of a case where a semantic difference really is going to matter to the behavior at hand. I can imagine plenty of cases where that semantic difference is going to matter somewhere down the line, but I also am inclined to think that in cases where that difference is going to happen down the line, I'm going to be able to find an internal difference that's going to explain it as well. I don't think that this will require me to go out into the environment to find the key proximal cause of the behavior. There is a sense in which these pictures of extended mind or situated cognition or embodied cognition are still relying on metaphors of the mind radiating out into the environment. These pictures are just very hard to reconcile with our existing neuroscientific data about how we go about explaining behavior.

Could you agree with someone saying: "OK, my account rides piggyback on your account when it comes to explaining bodily movements, but if you individuate actions more finely than that you have to bring in externally individuated mental factors"?

Well, I want to insist that it's still an open question empirically. We don't know the answer to it. For all we know the best explanations for certain kinds of phenomena may require intentional characterizations of not only the state you're using to do the explaining, but also the very behavior that you're explaining. I've never written a paper on this but I have often thought about writing a paper entitled "What the reductionist requires in order to pull off his or her program". It requires not only characterizing the states that are doing the explanation reductionistically, *but it also requires characterizing the behavior reductionistically*. It has always been my view that if you're a reductionist and you permit someone to describe behavior in ineliminably intentional terms, then you're in a game that you cannot win. When you provide the non-intentional causal-mechanistic explanation it's just going to seem unsatisfying, but if you're permitted to redescribe the behavior in non-intentional terms, your non-intentional mechanistic explanation is going to seem far more satisfying.

However, there is also the opposite worry. Suppose that we don’t care about the fine-grained intentional differences between behaviors just discussed. Rather we describe the explanandum roughly, for example, as “he walked through the door”. Then exactly which specific lower-level mechanism was leading to the effect might not be relevant for our explanatory concerns, and this gives reason to think that the best causal explanation will pick out some multiply realizable higher-level event or state as the cause.

Excellent point. And now we really are touching upon the way that debates about multiple realizability have moved on beyond where they were left off in the 1980’s and 1990’s. There is a reply to multiple realizability that is clearly formulated by Larry Shapiro, Mark Couch and Tom Polger.[8] Functionalists were allowed to assume identity of kinds at the mental level and then argue for very subtle differences at the physical level constituting multiple realizability and thereby an argument against reducibility. However, they were almost never challenged on this identification of kinds at the mental level. Moreover, as has been increasingly pointed out in the recent multiple realizability literature, the more similar in kind that you are at higher level of description, almost inevitably the more similar you are in terms of physical mechanisms. Only when there are great differences at the higher level of description then you tend to find these vast differences in physical mechanism. The reductionist should be permitted to say it’s not just the matter of you walking through the door, which of course admits of a wide variety of different kinds of behavioral realization, but of walking through the door at this particular speed through this particular pathway.

For genuine scientific purposes there’s almost no use for those broadly generic descriptions. For example, there is almost no use for the description of the eye as just a mechanism for seeing visual features of the world. Suppose you think that octopus eyes, mammal eyes and fly eyes are all of the same kind. They are all sensitive to photon information within a particular wavelength, but they are vastly multiply realized at the physical mechanism level. The problem with that view is that it’s very easy to find *higher-level* differences between these different kinds of eyes. You can easily find differences in the input-output features of these three different kinds of eyes. There is not a single kind of eye realized in these three different ways where that single kind eye figures in any serious science. As soon as you start insisting that people give kinds that figure in serious scientific examples it’s usually very easy to find behavioral differences, higher-level differences, when those kinds are realized in different physical mechanisms.

So the claim is that the scientifically interesting level of fine-grainedness favors the reductionistic approach?

Yes. Bill Bechtel and Jennifer Mundale developed the notion of grainedness in their article against multiple realizability.[9] Their point is that one of the reasons multiple realizability has looked so obvious is because the psychological stuff has been described very coarse-grained, while the neurobiological stuff has been described very fine-grained. However, the psychological stuff could equally be defined as fine-grained within psychology, and then we’ll probably find one-to-one matches between the fine-grained psychological descriptions and the fine-grained neural descriptions. Moreover, this is not just a conceptual point. In their article they actually try to show some evidence, mainly from neuroanatomy, that this point also is *empirically* justifiable.

When one focuses solely on the different cellular and molecular mechanisms, do you see any reason that a no-person worry might emerge: That the mechanisms are all acting in concert to produce the behavior, but that the person as a unified object doesn’t serve any explanatory role over and above this fragmented bunch of cellular and molecular mechanisms?

That’s a fascinating question. The concept of a person has so many different kinds of uses, for example in moral discourse, in political discourse and in social discourse. One of the points I was trying to stress at this conference is that your particular view about what the nature of human behavior is does not necessarily lead you to a particular view about the way that persons should be treated in political contexts. When it comes to political situations, I’m inclined to be completely pragmatic, and I don’t really much care if the notion of person that I find useful for thinking about how politics should be organized really matches up with my scientifically inspired notion of a person that comes from my attempt to analyze our best causal mechanistic picture of the world. I think too many people have thought that the only way you can be a political libertarian is if you couple that with metaphysical libertarianism, but there is no logical connection. I mean, there’s a logical connection in the sense that if you’re a metaphysical libertarian you probably are going to be some kind of political libertarian. However, the other direction doesn’t go. You can be a political libertarian on completely pragmatic grounds because you distrust the ability of human beings to apply scientific knowledge to political organization in a way that's going to benefit human beings. I don't think people are really free or autonomous in the metaphysical sense, but I just don't see any better way to organize a political structure other than treating them that way. And I don't personally see an internal conflict in that view.

I have to admit that this combination of views sounds almost like self-deception in my ears. Speaking normatively, there's a strong intuition that you shouldn't treat people as something they are not, doesn't this constitute a serious worry?

I freely admit to, like you, having that intuition. And it's a hard one to overcome. But I think what really causes me to overcome the pull of that view is to look at the kinds of political organizations that people who have adopted a "scientific perspective" on human beings have proposed. Personally, they're not the kinds of political organizations that I'd want to live in. Even if my picture of myself is radically false – even if there is nothing in the real world that really answers to autonomy and freedom and self-respect and all that – nevertheless I'm not willing to sacrifice those particular notions to live in some of these structures that have been proposed by my fellow hard determinists. Even if it's ultimately not based on any kind of metaphysical or scientific fact, I much prefer to live in a situation where I'm treated as a free autonomous agent that is responsible for my actions and is not submissive to government power over individual choices except in limited contexts. That's my choice of the kind of political and social organization that I find much more conducive to doing the kinds of things that I like to do, even if, in the end, there is nothing in our best causal mechanistic story of human beings that answers to that. In short, I think that we can be completely determined to think of ourselves as free autonomous agents.

That's a nice line to end with. Thank you very much Professor John Bickle.

NOTER

1 Ken Schaffner is Professor of History and Philosophy of Science at the University of Pittsburgh.

2 Jaegwon Kim is Professor of Philosophy at Brown University. His book *Mind in a Physical World* (MIT Press 1998) is perhaps the most important contribution to the mental causation debate in the last two decades.

3 Johan Storm is Professor at the University of Oslo.

4 Ernest Nagel's account of scientific reductionism has been highly influential. It was presented in his 1961 book *The Structure of Science*.

5 Multiple realization is the idea that the very same mental state can be realized by different physical states.

6 Robert Richardson is Professor of Philosophy at the University of Cincinnati.

7 Bickle refers to the so-called Morris water maze test. It was designed by Richard G. Morris, and is used in research on spatial memory.

8 Larry Shapiro is Professor of Philosophy at the University of Wisconsin, Madison, Mark Couch is Assistant Professor of Philosophy at Seton Hall University, and Thomas W. Polger is Associate Professor of Philosophy at the University of Cincinnati.

9 "Multiple Realizability Revisited: Linking Cognitive and Neural States" in *Philosophy of Science* Vol.66, 1999, page 175-207. Jennifer Mundale is Associate Professor of Philosophy at the University of Central Florida, and Bill Bechtel is Professor of Philosophy at the University of California San Diego.