Resnick, M. (1994b). Learning about life. *Artificial Life*, *1*, 229–242.
Ross, D. (2004). Meta-linguistic signalling for coordination amongst social agents. *Language Sciences*, *26*, 621–642.
Saltzman, E. (1995). Dynamics and co-ordinate systems in skilled sensorimotor activity. In R. Port & T. van Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (pp. 149–173). Cambridge, MA: MIT Press.
Steels, L. (1994). The artificial life roots of artificial intelligence. *Artificial Life*, *1*, 75–110.
van Gelder, T. (1991). Connectionism and dynamical explanation. *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 499–503). Hillsdale, NJ: Erlbaum.
Wheeler, M. (1994). From activation to activity. *Artificial Intelligence and the Simulation of Behaviour (AISB) Quarterly*, *87*, 36–42.
Wheeler, M. & Clark, A. (1999). Genic representation: Reconciling content and causal complexity. *British Journal for the Philosophy of Science*, *50*, 103–135.
Wimsatt, W. C. (1986). Forms of aggregativity. In A. Donagan, N. Perovich, & M. Wedin (Eds.), *Human Nature and Natural Knowledge* (pp. 259–293). Dordrecht: Reidel.
Wimsatt, W. C. (2000). Emergence as non-aggregativity and the biases of reductionism(s), *Foundations of Science*, *5*, 269–297.

# 11

# IS PSYCHOLOGICAL EXPLANATION BECOMING EXTINCT?

*Cory D. Wright*

## 1. Intertheoretic Reduction versus Reductive Explanation

The doctrine of psychoneural reductionism has long been construed as ultimately rooted in traditional issues in metaphysics and philosophy of science; this is perhaps no more manifest than in attempts to reformulate the mind/body problem as a special application of intertheoretic relations between formally structured pairwise theories replete with lawlike generalizations. Construed thus, proponents of the doctrine sharply distinguished *theory reduction* from *reductive explanation* (Achinstein, 1984; Brandon, 1984; Churchland, 1986; Bickle, 1998, 2003; Barendregt & van Rappard, 2004; Wright & Bechtel, 2006), leaving them to restrict their efforts to the rational reconstruction of these idealized or "textbook" relations using formal set-theoretic analyses. Reductive explanation, on the other hand, could be shuffled off onto historians and sociologists of science, or those whose interests led them to dwell on social, pragmatic, or purely methodological issues involved in the investigation of some target phenomenon Φ. After all, "reductive explanation" – like "explanation" more generally – designates a *communicative practice* between groups of cognizers (be it a butcher, baker, or candlestick maker (or neuropharmacologists)).[1] And mulling over the personal habits and discursive quirks of various researchers hardly seemed a task worthy of the metaphysician or philosopher of science.

The distinction between theory reduction and reductive explanation – paralleling the old "logic of justification" versus "discovery" distinction – afforded psychoneural reductionists a means of compartmentalizing irreducibility claims from the "antireductionist consensus": the *informal explanatory patterns and practices* for some Φ might be carried out independently of lower-level research as long as that independence ultimately didn't carry over to *formal intertheoretic* contexts *as irreducibility*. Such autonomy/independence wasn't threatening to intertheoretic reductionists, since autonomous methods and explanatory practices are crucial to scientific progress, and since the methods and day-to-day explanations inherent in scientific research aren't exhaustively reductive anyway. Bickle, following Churchland (1986, p. 381), effectively makes this type of reply to Pylyshyn:

> One can predict an intertheoretic reduction without tying one's methodological practices to reductive explanations. Reductive *explanations* seek to explain *phenomena* described in a higher-level vocabulary as explananda by exclusive appeal to lower-level kinds and laws as explanans. . . . An inter*theoretic* reductionist can agree wholeheartedly with this methodological point. He need have no commitment to the exclusive use of reductive explanation. . . . Reduction is a prediction about how candidate *theories* will comport with one another *after the fact* of maturity and development. (1998, pp. 153–154)

With this distinction, psychoneural reductionists could keep irreducibility claims at bay in cases where mature or developed theories are reductively related, whilst making the harmless concession that higher-level psychological research has a "useful and seemingly ineliminable heuristic role to play" (Churchland, 1986; Bickle, 1998, p. 153; 2003, p. 130; Craver, 2001).[2] Both in-principle irreducibility of pairwise theories *and* the exclusive use and import of reductive explanatory practice could therefore be rejected *without* inviting charges of inconsistency.

I think that this concession is both sensible and warranted, for broadly Feyerabendian reasons: intertheoretic contexts aside, explanatory practice is hardly *exhausted* by decomposition or other reductive heuristics – crucial and necessary though they are (Wright, 2002). For Bickle, Kim, and other reductionists, this concession apparently conceded too much.[3] Explanatory autonomy/independence of reductive explanation are now also deemed impermissible: for any given Φ,



when sufficient amounts of lower-level achievements allow for a successful reductive explanation to obtain, further explanatory contributions from higher-level research are thereby *precluded* because they aren't able to continue offering genuinely penetrating causal insights about the phenomena within their scope. The effect of preclusion on higher-level explanations issuing from such research would seem to be *extinction*. Call this supposition "explanatory extinction" (*EE*).[4] In this chapter, I argue that appeals to actual scientific practice show that higher-level psychological explanations aren't precluded by successful reductive explanation, much less rendered extinct.

## 2. Psychological Explanation as Fossil Record?

To be clear, *EE* is a supposition about the consequences of successful reductive explanations on higher-level explanatory patterns and practices. *EE* says that the various discursive and explanations typifying the psychological sciences (e.g., adaptive, cognitive, computational, functional) are *de facto* expunged from ongoing research once they discharge the task of enabling increasingly accurate neurobiological explanations; the advent of accomplished finer-grain stories will sanction the act of jettisoning the then-superfluous higher-level stories, like the scaffolding on a finished building. Of course, reductionists can certainly allow, expect, and welcome results from cognitive and clinical neuroscience, experimental psychology, and phenomenology that enable and guide lower-level sciences by, e.g., helping to locate and determine the bounds of the *explanandum*, supplying their own data about higher-level concepts and categories, and providing the necessary initial approximations. In that sense, reductionists have no truck with principles of mutual co-evolutionary feedback and level-bridging; but once a successful reductive explanation obtains, the psychological explanations involved are marked for extinction (Schweizer, 2001; Bickle, 2003, 2006).

There's perhaps no more able a reductionist than Bickle to cite with respect to *EE*, who is worth quoting at length on the matter:

> [W]hen we fix our gaze on aspects of scientific practice in [actual case studies], we see that psychological explanations lose their initial status

> as causally-mechanically explanatory vis-à-vis an accomplished cellular/molecular explanation . . . Within scientific practice, psychological explanations become otiose when the cellular/molecular explanation . . . is achieved. There is no need to evoke psychological causal explanations, and in fact scientists stop evoking and developing them once real neurobiological explanations are on offer. Philosophers who deny this point are usually guided by outdated accounts of real neuroscientific practice. Contra Kim, lower-level explanations need not "exclude" higher-level accounts in any deep epistemological or metaphysical sense. But the former do render the latter pointless, along with any further search for empirically improved successors at the same level – except for some residual, purely heuristic tasks. [A]ccomplished lower-level mechanistic explanations absolve us of the need in science to talk causally, or investigate further, at higher levels, at least in any robust "autonomous" sense. To articulate and defend these claims, I propose that we let scientific practice be our guide. In light of this existing cellular/molecular explanation and these experimental results, it seems silly to count psychology's "explanation" of consolidation as "causally explanatory", "mechanistic", or a viable part of any current scientific investigation worth pursuing. (2003, pp. 110–112)

Plainly, Bickle here asserts *EE*, and avers that failing to do so results from spending too much time gawking at fossil records. The idea, then, is straightforward. When successful reductive explanations obtain, the psychological explanations involved are rendered "pointless" or "silly" curios whose significance becomes entirely historical. The very achievement of successful reductive explanations demands that psychological explanations cede way; and once precluded, they inevitably lose – never to regain – their status as genuinely explanatory pieces of the causal story.

## 3. Bickle's Wager

If *EE* were true, it would certainly encourage the doctrine of psychoneural reductionism. But is it? And is there an argument for it's being true? What might such an argument look like? In Bickle's assertion of *EE*, the basis for thinking that it's true is not so much an argument as it is a *wager* that successful reductive explanations inevitably *infantilize* higher-level explanatory contributions. The reductionist



bets that, by attending to current scientific practice, one will *just see that* such contributions are precluded whereupon a successful reductive explanation obtains.

For the sake of a name, call this wager "Bickle's wager." (To be clear, the target of criticism is neither Bickle himself nor his versions of psychoneural reductionism, but reductionists' deployment of what I am calling "Bickle's wager" to establish the truth of *EE*.) Ironically, Bickle's wager embodies the old Wittgensteinian injunction, "Look, don't think!" Its characteristic devil-in-the-details approach – which fixates on the so-called "put up or shut up" challenges that have recently populated the reductionism literature – proposes that *mere attention to* experimental results in neuroscience will likely reveal accomplished reductive explanations. Philosophers need only look at what's being done in labs across the world, because what's being done just is the reductive explanation of psychological phenomena in terms of explanations of phenomena at much lower levels.

True to form, Bickle antes up with an impressive – though by no means uncontroversial – case study (Schouten & Looren de Jong, 1999; Bickle, 2003, 2006; Craver, 2003). Higher-level psychological explanations of memory, which appeal to the conversion of short-term memory traces to stable long-term memories through stimulus repetition and in the absence of retrograde interference, have been extinguished because of successful reductive explanations of the memory consolidation switch in associative learning theories to the complex mechanisms producing early- and late-phase long-term potentiation (E- and L-LTP). Neuroscientists have shown that this functionally construed posit is subserved by a series of mechanistic activities in CA1 and CA3 regions of the hippocampus (HC). Hence, what was previously defined as a "consolidation switch" at the level of psychological function is now understood as increased synaptic potentiation at cellular/molecular levels; the production of robust excitatory post-synaptic potentials (EPSPs) in mnemonically dedicated circuits results from increased gene expression – specifically, the transcription of genes expressing CREB proteins. Increased gene expression, in turn, results from the activation of protein kinase A and the release of its subunits (which turns off the inhibition of phosphorylated calcium-calmodulin kinase II (CaMKII)), and the conversion of second messenger systems such as cyclic adenosine monophospate (cAMP). For reductionists making Bickle's wager, the larger point is that (putatively) successful

reductive explanations like this one can be compiled and parlayed into a strong induction for the doctrine of psychoneural reductionism.

Under some descriptions, this devil-in-the-details fixation is surely worthy of applause. Put innocently enough, there's no good reason not to concur: e.g., "Some might find it difficult in the abstract to see how mind reduces to molecules; but one will never get over that intellectual hurdle or show conclusively why such reductions can't obtain by remaining ignorant of the best existing scientific attempts to do exactly this" (Bickle, 2003, p. 95). But its more extreme forms have nontrivial consequences that should distress philosophers of psychology or cognitive science. For it relegates their role to that of being little more than "science journalists" – i.e., merely reacting to late-breaking neuroscientific developments, and then rehearsing whatever inferences about the mind neuroscientists happen to present them with (albeit in a more popular vernacular). In such extreme forms, the philosophy in "neurophilosophy" is sloughed off along with like-minded disciplines. Consider, e.g., recommendations to bypass research appearing in journals such as *Brain and Behavioral Sciences*, *Journal of Cognitive Neuroscience*, and *Psychological Review* (Bickle, 2006); such eliminativist catcalls portend a revision to the relevant bit of Hume's *Inquiry Concerning Human Understanding*: "Does it contain any (cellular/molecular neuroscientific) experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames, for it can contain nothing but sophistry and illusion."

Bickle's wager requires reviewing experimental results from the best of what science has to offer. So, in Section 4, I'll do just that. I'll review an exemplary case study from behavioral neuroscience – namely, the reductive explanation of an aspect of reward function in terms of dopaminergic operations of the mesocorticolimbic system. If Bickle's wager were a good bet, this successful reductive explanation would establish *EE*. But it doesn't, and *modus tollens* gives you the rest: psychological explanation isn't in danger of going extinct.

## 4. Mediation of Incentive Salience by Mesocorticolimbic Dopamine

The concept of REWARD falls at the heart of a class of higher-level functional and intentional constructs (desire, appetitive/consummatory



pleasure, craving, arousal, reinforced learning, etc.). At least two main features structure this class. First, these constructs are said to possess *direction*, insofar as they are involved in the selection, initiation, and performance of behaviors toward the achievement of some end or goal. Second, they are said to possess *hedonic valence*, insofar as the states and processes they pick out are linked to positive/negative affective values. Reward function exemplifies both of these features; it makes available to an animal a basic functional capacity to be attracted to certain stimuli and to move on the basis of such attraction, and its dysfunction involves characteristic abnormalities in attraction. And so, like most all constructs in this class, it concerns the relationship between an animal's neuropsychobiological profile and its behavior, i.e., those states and processes responsible for its goal-directed commerce in specific environments. Obviously, much more needs to be said about REWARD, what it represents, what a reward function is and how it interacts with other capacities like attention or motivation, whether REWARD can be accurately characterized by causal role functionalism, etc. (In some sense, answers to these questions are precisely what psychoneural reductionists forego; I'll return to this point again in Section 6.)

Fifty years of converging evidence – from a range of species, particularly rats – indicates that the intersection of the sets of neural circuits in which reward function is realized involves a common substrate in the mesocorticolimbic system. Generally, this system – when conjoined with the mesostriatal system – tightly "binds" the directive and hedonic capacities of motivation and pleasure with motor abilities for ascertainment behavior. And dopamine (DA) is one of the main components. Around 80 percent of DA neurons are synchronically activated in mechanisms producing reward (Schultz, 1998), and pharmacological blockade with DA antagonists is well known to induce reward-functional impairments. Within the mesocorticolimbic system, the connectivity between the ventral tegmental area (VTA) and nucleus accumbens (NAc) – 85 percent of which is composed of DA neurons – has garnered much attention as the most significant pathway-governing reward function (Heimer et al., 1991; Fiorino et al., 1993; Pfaus et al., 1995; Ikemoto & Panksepp, 1999; Salamone & Correa, 2002; Hjelmstad, 2004; Wise, 2004).

But how, exactly, do the operations of DA neurons that factor in the activities of the mesocorticolimbic system govern reward functional

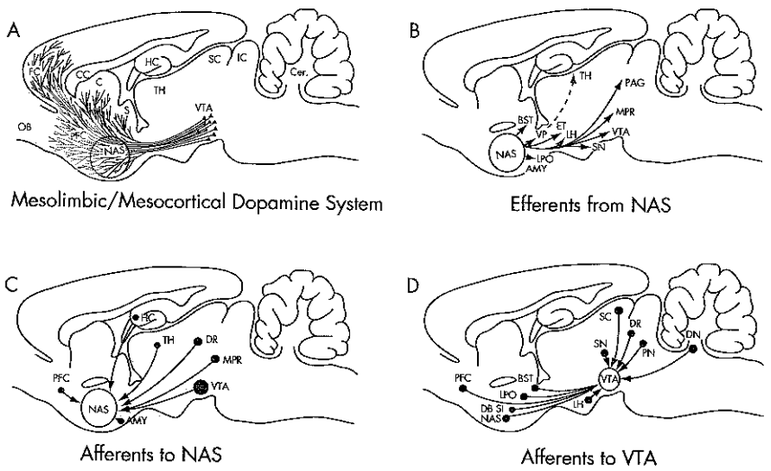


**Figure 11.1** Schematic diagrams of mesocorticolimbic pathways in the rat brain
*Note*: NAS = Nucleus Accumbens, VTA = ventral tegmental area, LH = lateral hypothalamus, and HC = hippocampus.
*Source*: Reprinted from S. Ikemoto and J. Panksepp, "The role of nucleus accumbens dopamine in motivated behavior," *Brain Research Reviews, 31* (1999): 9, with permission from Elsevier.

states? A now-standard explanation proposes that rewards preferentially potentiate DA transmission in the shell of the NAc (Pfaus et al., 1995; Wise, 1998, 2004; Di Chiara, 1999; Spanagel & Weiss, 1999). Under baseline conditions, DA neurons tonically emit action potentials at low frequencies (≈5–10 Hz), but emit short and sudden *phasic bursts* of high-frequency activity (≈15–20 Hz) that become more prevalent in response to rewarding stimuli – especially those that are *salient* or *novel* (Freeman, Meltzer, & Bunney, 1985; Schultz, 1998; 2002; Casassus, Blanchet, & Mulle, 2005). Functionally, this requires the joint coordination of additional perceptual, attentional, and motivational activities. For example, Horvitz, Stewart, & Jacobs (1997) monitored isolated EPSP waveforms of feline VTA DA neurons in response to visual and auditory stimuli, and discovered that the mere presentation of stimuli dramatically increased the probability of phasic bursting within 200 ms and returned to tonic baseline firing immediately thereafter. By nonlinearly increasing these burst rates, DA neurons



can flood a synaptic cleft with "transients" – rapid surges in extracellular DA concentrations that are unachievable in the same time frame with just the same number of normal, low-frequency activation potentials. The failure of expected rewards results in activation patterns that fall back to single spikes (Schultz, 2002). VTA DA neurons, then, provide signals for distinguishing between non-/rewarding stimuli by generating and altering activation patterns induced by phasic bursting, which increase extracellular DA release at the NAc; as they do so indiscriminately across various types of natural and artificial rewards regardless of sensory modality, this operation seems to be the primary neurochemical event that is causally responsible for many aspects of reward function and positive reinforcement (Kiyatkin, 1995).

DA's role in reward function has proven rather more subtle: converging evidence provides reason for thinking that DA doesn't directly code for all aspects thereof, and that many other components and operations are involved (Berridge & Robinson, 1998; Ikemoto & Panksepp, 1999; Spanagel & Weiss, 1999; Berridge, 2003; Wise, 2004; Salamone et al., 2005). Using a DA reuptake blocker and $DA_1$ receptor antagonist in a probabilistic nosepoke design, Nicola et al. (2005) demonstrated that NAc DA release is both necessary and sufficient for eliciting the causal cascade that eventually produces approach behaviors to rewarding stimuli. Hence, the integrity of the mesocorticolimbic and mesostriatal systems is extremely crucial for being *attracted* by rewarding stimuli; but is DA crucial for coding sheer qualitative pleasure, or in consuming rewarding stimuli? Increases in transient concentrations have been demonstrated by monitoring extracellular changes in DA and DA metabolites (e.g., DOPAC) using *in vivo* microdialysis studies during intracranial self-stimulation of the lateral hypothalamus (LH) (Fiorino et al., 1993; Di Ciano, Blaha, & Phillips, 1998). Self-stimulation is often interpreted as being rewarding because it deceptively activates mesocorticolimbic DA systems that normally code for attractive, novel, or salient stimuli. Yet, Neill, Fenton, & Justice (2002) measured DA and DOPAC concentrations in different LH self-stimulation designs, and found that increased DA firing in the NAc was primarily influenced by the schedules of reinforcement, leading them to conjecture that DA release modulates the response effort and perceived cost of behavior rather than the sheer qualitative pleasure or the actual hedonic properties of stimulation themselves (see Ikemoto & Panksepp, 1996).

Another, more prominent example of subtle refinements to the reductive explanation comes from Berridge and colleagues' studies of unconditioned affective behaviors to food rewards. Based on taxonomies of homologous types of animals' facial expressions and affective behaviors to a variety of tastes and food rewards, they developed an experimental procedure that ostensibly yields observable indications of hedonic valence following manipulations of cellular/molecular DA operations (Berridge, 2000, 2003; Berridge & Robinson, 1998; Robinson & Berridge, 2000, 2004). In one experiment, they induced massive bilateral lesions (>98 percent) of ascending DA neurons from the VTA and DA terminals in the LH and NAc of rats with direct injections of 6-OHDA neurotoxin. If DA operations directly mediated hedonic valence, such total depletions of DA in these key areas should have nullified the ability of rats to experience pleasurable rewards; but neither unconditioned affective behaviors to the positive hedonic valence of sucrose solution nor to the negative hedonic valence of quinine HCl were impaired. (The researchers deduced that this effect wasn't the result of disrupted organization of forebrain and mesocortical systems by the DA lesions, which would be needed for intact associative processes for evaluating the rewarding properties of sweet rewards.) Further, since benzodiazepine agonists enhance positive hedonic valence, then if DA-depleted rats don't suffer from deficits in valence, they should have heightened affective behaviors to sucrose; to confirm this conclusion, diazepam was administered to lesioned rats, which subsequently *did* show the heightened affective behaviors. Accordingly, Berridge and colleagues concluded that mesocorticolimbic DA operations aren't necessary for normal hedonic evaluations (Berridge & Robinson, 1998, p. 328).

Of interest here is that Berridge and colleagues were led to invoke a higher-level distinction to dissociate between aspects of reward function – what they term "core liking" versus "core wanting" – in order explain that two neural subsystems are involved in mediating two different psychological processes. "Core liking" – which was shown to be mediated by a subsystem of which DA isn't a primary component – refers to the attribution of positive hedonic valence to representations of rewarding stimuli (basically, the process of taking some reward to be pleasurable), and is dissociated from "core wanting," which refers to the subpersonal psychological process of forming and



transforming representations of ordinary stimuli such that they become especially salient, "attention-grabbing," or otherwise attractive. This distinction – which can be summarized simply as: "one need not want what they like or like what they want" – reformulates a pair of old folk-psychological notions to help make sense of the functional roles played by mesocorticolimbic DA operations.

The latter transformational process, mediated by VTA DA phasic bursting, is necessary for constructing and processing neural representations that attribute *incentive salience* to rewarding stimuli (Berridge & Robinson, 1998; Ikemoto & Panksepp, 1999; Wise, 2002, 2004; McClure, Daw, & Montague, 2003; Robinson & Berridge, 2004; Salamone et al., 2005). Incentively-salient neural representations serve as intermediary links between the activation patterns induced by phasic bursting of VTA DA neurons at NAc terminal fields, and the higher-level cognitive representations interfacing with various reward functional states more generally. These include other attentional and perceptual, affective, learning and memory, motivational, and motor processes that jointly constitute an animal's abilities to ultimately recognize positive reinforcers *as* having a certain hedonic valence, and to elicit approach and even long-term, goal-directed behaviors. Neural representations of incentive salience – and the DA operations that mediate them – thus have a crucial place within the hierarchy of other states and processes at various levels, which constitutes the mesocorticolimbic system and the final common pathway for a repertoire of flexible approach behaviors.[5]

## 5. A Rewarding Case of Reductive Explanation?

In formal contexts of theory reduction, psychoneural reductionists have been unable to determine how to precisely locate certain cases on the intertheoretic reduction spectrum (Wright, 2000; Fonseca, 2004). For instance, Churchland concludes that determining when a perfectly smooth reduction obtains is a matter of "the whim of central investigators, the degree to which confusion will result . . . , the opportunities for publicizing [results], cadging grants, and attracting disciples" (1986, pp. 283–284); likewise, Bickle concedes that it's

"ridiculously optimistic" to provide formal criteria for distinguishing between perfectly bumpy reductions from eliminations (Bickle, 1998, p. 101). In informal contexts of reductive explanation – where success and sufficient completeness of investigation and understanding is much more nebulous – providing positive criteria is all the more difficult; unfortunately, asserting *EE* incurs the onus of doing just that.

Since it's unlikely that reductionists will have such criteria in informal contexts of reductive explanation, let's simply grant them, for the sake of argument, that the reward function case is a case of successful reductive explanation. This is an eminently plausible assumption. For what I've presented barely scratches the surface of a massive corpus of research on the realization of aspects of reward function. And there's no doubt that epistemic progress has partially resulted from lower-level explanatory pressures, like the resolution of important anomalies. For example, one problem for past explanations of reward function in terms of mesocorticolimbic mechanisms is that well over two-thirds of relevant DA neurons seem to be excited by aversive stimuli (Ikemoto & Panksepp, 1999). Recent electrophysiology and immunofluorescence work by Ungless, Magill, & Bolam (2004) demonstrated that the anomaly is only apparent: DA neurons *are* uniformly inhibited by aversive stimuli because the aversively-excitable neurons weren't actually dopaminergic. Many further refinements still need to be made, details specified, anomalies resolved, more penetrating formulation of constructs, etc.; but even this all-too-brief review exhibits the stable outlines of a successful reductive explanation already hewn. At the very least, psychoneural reductionists would be hard-pressed to deny it. For one thing, the reductive explanation of incentive salience in terms of DA bursting/gating easily rivals that of memory consolidation in terms of E- and L-LTP; hence, denial would therefore put reductionists in the uncomfortable position of seeming hasty or disingenuous about the success of the latter case. Moreover, denial would fail to square with many reductionists' staunch dismissal of epistemic impoverishment claims (i.e., bemoanings of "how little we know about how the brain works"; Bickle, 2003, 94–95); hence, if such claims are indeed gross miscalculations about what is known, then surely this case constitutes a successful reductive explanation if any does (i.e., it's hard to feel sympathetic with the reductionist who tries to have it both ways).



## 6. The Importance of Being Level-Headed

It's important to note that even successful reductive explanations are susceptible to explanatory refinement: there's always the possibility of compiling more information, of better stories to tell, different ways of telling them, and new connections to forge – even "after the fact of maturity and development." Intuitively, it might appear that susceptibility to explanatory refinement entails that the reductive explanation wasn't actually an accomplished case after all. But this intuition implies that reductive explanations need to be complete to count as successful; and surely such a requirement is too strong, since most reductive explanations would thereby be rendered "unsuccessful." In any case, the appearance of inconsistency in psychoneural reductionist doctrine – particularly, in Bickle's wager – easily dissipates *if* explanatory refinements aren't taken to originate from higher-level research. For recall that anyone making Bickle's wager is only committed to a conditional: if we attend to the details of actual scientific practice, we will find that *EE* holds in cases where accomplished reductive explanations obtain. So consistency dictates only that reductionists assert the further claim that refinements are permissible if they take a "more-of-the-same" form (e.g., making more low-level manipulations and specifying the causal effects on higher-level psychological processes and behavior); hence, *once* a successful reductive explanation obtains, only higher-level research loses its explanatory and revisionary potency. It's this further assertion that should be contested. For starters, if dissolving the appearance of inconsistency requires that explanatory refinements don't originate from higher-level psychological research, then *EE* is a patently question-begging supposition, since that's the very issue under dispute.

The "intervene-cellularly/molecularly, track-behaviorally" strategy purportedly best captures experimental and explanatory practice (Bickle, 2003, 2006); and indeed, reward researchers often utilize, e.g., lesion/deficit preparations to understand which components' absence disrupts normal mechanistic activity, as well as preparations involving excitation and inhibition of lower-level component operations. Lesion experiments with 6-OHDA neurotoxin are a prime example. But while reductionists are quite right to focus on reductive explanations born of these sorts of reductive procedures, the "intervene-and-track"

strategy is hardly the only experimental procedure used by researchers to further develop already accomplished explanations. Many researchers also use what are called "contextual" or "additive" strategies (Craver, 2001), such as methods that track the effects of environmental conditions or cues on behavioral paradigms, or that investigate the effects that larger, more encompassing mechanisms have on the component parts in which they operate.[6] For instance, a crucial means of implicating phasic bursting of DA neurons in incentive salience and reward-related learning was to simply chronicle what individual DA neurons do when the animal as a whole is put in a particular environment and presented with different types of distal stimuli (Horvitz et al., 1997). In another single-unit recording study, Taha and Fields (2005) implanted electrode arrays in rats' NAc and then ran the animals through two behavioral paradigms – a sucrose discrimination task and a contrast task to determine whether NAc cells code for hedonic valence. By recording the activities of select neurons and then manipulating the whole animal and its environment, researchers were able to distinguish two operations performed by the neural population: excitations encoding the hedonic valence of sucrose solutions, and disinhibitions of other neurons to initiate and maintain consummatory behaviors. As a matter of public record, these non-reductive, "top-down" strategies provide refinements even *after* successful reductive explanations has obtained; ergo, rather than merely enabling more lower-level explanation, they provide new information based on manipulations of higher-level variables. The new information provided is usually of the sort left out of reductive explanations – namely, information about how functions and mechanistic activities are embedded in more encompassing systems, about functional significance and better role specifications, about the effects of environmental or adaptive contexts in which Φ is situated, better understanding of systemic-level activities, and so forth. After all, VTA DA neurons must be causally hooked up to each other in appropriate ways in order to provide signals that form and transform incentively-salient neural representations. And the circuits they compose must themselves be organized so as to efficiently interact with larger and more complex mechanistic subsystems, which are themselves embedded within larger mesocortical, mesolimbic, and mesostriatal systems. But as increasingly sophisticated explanations then circle back to explain these higher levels of systemic organization – such that the ultimate



set of relations between *explanans* and *explanandum* can be understood – the significance of each individual lower-level component cedes explanatory priority (Wright & Bechtel, 2006, p. 61). With this "circling back," the explanatory focus becomes the overall activity of the larger circuits and mechanistic systems comprising the final common pathways affording cognizers the very abilities and capacities to construct nuanced cognitive representations of reward-related learning, behavior, and stimuli.

Of course, contextual and additive strategies – while casting doubt on reductionists' claims that explanatory refinements only originate from lower-level research – aren't yet enough to save psychological explanation from extinction. It must be shown that psychological-level research itself still plays some role, however minimal. Here, examples are more difficult to find – but not impossible. One way that psychological experiments are crucial is in helping evaluate experimental procedures and models against certain kinds of validation criteria – particularly, convergent and discriminant validity, and face validity (Wright, 2002). For instance, Germans and Kring (2000) show that some self-report measures of anhedonia (e.g., Scale for Physical Anhedonia) aren't adequately sensitive to detect differences in affect, but that other measures – such as self-report of approach motivation and motivational salience – may better index hedonic deficits. Such (purely) psychological studies corroborate what is known about the general role of motivation and incentive salience in reward function and provides reason to filter out particular clinical measures that may be obscuring what little human or clinical data is available. The interesting thing about this sort of example is that, *despite* giving an extremely anemic sense of what higher-level psychological research on reward function amounts to, it's nevertheless *strong enough* to show why an understanding of the cellular/molecular mechanisms of reward function, once achieved, doesn't simply leave psychological explanations with *nothing* to do.[7]

Another reason for contesting the question-begging assertion that explanatory refinements can't originate from higher-level psychological research is that successful reductive explanations don't simply invoke successively lower-level vocabularies. Communicative practices being what they are, scientific explanations are often constructed using miscegenated vocabularies from different subdisciplines – there simply are no neat divisions among lexical items. This should not be

taken lightly. Consider, e.g., the reward function research that's been taken up by the burgeoning field of neuroeconomics (Schultz, 2004). Bridging neuroscientific and economic explanations would hardly be possible were it not for the higher-level psychological constructs involved in explaining reward-related learning, expectation, uncertainty, and decision making. DA operations in reward-related learning, for instance, involve numerous appeals to information-theoretic and cognitive terms ("processes feedback information," "signal," "uncertainty," "evaluating expectations," etc.). Hence, it should be no surprise why the reductive explanation of this aspect of reward functionality nontrivially interprets mesocorticolimbic DA operations using the conceptual apparatus (e.g., PREDICTION, ERROR-RECOGNITION, EXPECTATION) of higher-level information theory (Gallistel, 1994; Cohen, Braver, & Brown, 2002; Schultz, 2004; Casassus et al., 2005). For example, Schultz, Dayan, & Montague (1997; Schultz 1998, 2002) used computational models of temporal difference-learning algorithms to demonstrate that DA firing patterns in the VTA and substantia nigra (SN) obey formulas representing the deviation or error between predictions about the timing, direction, and valence of rewards, and the properties of those rewards which actually obtain. These discriminatory predictions of uncertainty and error vis-à-vis expected hedonic valence are what we might call "protoepistemic processes," and are themselves organized and coordinated such that they factor into larger, more complicated epistemic or cognitive processes. McClure et al. (2003) adapted these same computational models to show how the concept of EXPECTED FUTURE VALUE in Schultz & colleagues' work could be identified with the concept of INCENTIVE SALIENCE ATTRIBUTION in that of Berridge and colleagues'. The result of such bridging was "a more formal computational theory of how [DA] is involved in a larger system for choosing optimal actions under the motivation of prediction errors" (2003, 427; cf. Schweizer, 2001). Suffice it to say that successful reductive explanations invoke a variety of ideas and distinctions originating from numerous levels; Berridge and colleagues' employment of "liking" versus "wanting" to clarify the role of DA in incentive salience attribution nicely exemplifies this. As a blend of folk and scientific-psychological constructs, the distinction allowed researchers to interpret theoretically complex results in terms of a simpler conceptual domain that is already well understood. And as a slice of actual scientific practice, understanding a poorly understood



or unwieldy domain in terms of something more familiar or comprehensible is undoubtedly the norm in scientific reasoning. Successful reductive explanations are no different, synthesizing a variety of conceptual structures and relying on interpretations of data derived from models and experimental techniques that integrate phenomena of different grain and particularity.

This last point becomes more prominent still upon noting that neurobiologists themselves constantly make explicit calls for further evidence and constraints from the psychological sciences. In their review, Robbins and Everitt concluded, "Even leaving aside the complications of the subjective aspects of motivation and reward, it is probable that further advances in characterizing the neural mechanisms underlying these processes will depend on a better understanding of the *psychological* basis of goal-directed or instrumental behavior" (1996, p. 228; my emphasis). In particular, they noted an immediate need for functional neuroimaging results in cognitive neuropsychology to help localize large-scale task-relevant information-processing operations. Similarly, Berridge and Robinson aver, "[F]urther advances will require equal sophistication in parsing reward into its specific *psychological* components . . . Neuroscientists will find it useful to distinguish the *psychological* components of reward because understanding the role of brain molecules, neurons, and circuits requires understanding what brains really do – which is to mediate specific behavioral and psychological functions" (2003, p. 507; my emphasis). So, have reductionists shown that "The causal explanations of quantitative behavioral data eschew appeals to concepts at levels higher than cellular/molecular mechanisms and the anatomical circuitries containing them" (Bickle, 2006)? Not remotely. Indeed, the assertion of *EE* now appears awkwardly committed to the further supposition that neurobiologists themselves suffer from a sort of false consciousness or are otherwise deluded. For what's "explicit in scientists' writings" is the claim that better scientific understanding of what's been successfully explained demands *continued* conceptual contributions at higher levels.

The fact that neurobiologists make such calls for developments from psychological sciences *despite* having a wealth of lower-level data to invoke underscores the fact that making sense of the significance of phasic bursting of VTA DA neurons depends on how well illuminated the reward constructs involved are – i.e., the extent to which the

process of incentive salience attribution or "core wanting" is understood, what motivation processes are involved, what CRAVING is how appetitive behavior is different from consummatory behavior or approach responses, etc. Since the context here is explanation and not theory reduction, *understanding* is crucial. If the very concept of REWARD, or any of the fragmented constructs from that general class of phenomena, are undercharacterized or conceptually impoverished at the time of successful reductive explanation, then obviously there higher-level developments that still need to be made. This point is made plain by reconsidering how Section 4 began – with an intentionally all-too-brief account of the concept of REWARD, what a reward function is, whether the extensions of "reward" and "positive reinforcement" pick out equivalence classes of isomorphic properties, etc. In some sense, the brevity of such descriptions serves as a litmus test: even after a successful reductive explanation obtains, is further characterization of higher-level constructs needed? I see no good reason for thinking that researchers at higher levels should dispense with the desideratum of self-critically improving on their own psychological explanations. Using the construct of MOTIVATION as his example, here's how Berridge puts the point:

> Motivational concepts are becoming widely recognized as needed to help neuroscience models explain more than mere fragments of behavior. Yet, if our motivational concepts are seriously wrong, our quest for closer approximation to brain-behavior truths will be obstructed as much as if we had no concepts at all. We need motivational concepts, and we need the right ones, to properly understand how real brains generate real behavior. (2004, p. 180)[8]

So, even with a successful reductive explanation of aspects of reward function at hand, there's still a need, desire, craving, drive, urge, hunger, impulse, motivation, incentive, etc., for continued refinements in higher-level characterizations.

Accordingly, neurobiological developments notwithstanding, research on reward function is no different than virtually any other case study in the following respect: purely insular research that deliberately eschews bridges with research at other levels once accomplishments obtain is a rarity. Instead, investigations typically span several different levels – of both organization and analysis – and most experiments utilize various



techniques that integrate domains (Wright & Bechtel, 2006). If so, then it stands to reason that the explanations issuing from actual scientific practice cannot but help reflect this multilevel, pluralistic approach. A candid look at the discourse of actual scientific practice reveals why a broadly Feyerabendian analysis is gripping: rather than being restricted to research from increasingly lower levels, dis-/confirming supplementary results potentially come from any direction, from any level, from any branch of science or epistemology. Good explanatory practice just simply exploits that fact.

## 7. Conclusion

In order to make good on *EE*, the reductionist must demonstrate that actual scientific practice is played out in a certain reductionist key, and that philosophers need only attend to the details in order to hear the tune. I hope to have shown that the details don't quite pan out that way. The obtaining of a successful reductive explanation, born of accomplishments of lower-level research, fails to serve as a condition on the impossibility of achieving better scientific understandings of the target phenomenon $\Phi$ at higher levels. Now, reductions – when they occur – are fine so far as they go. Indeed, the construction of models involving the localization of mechanistic activities and the decomposition into their component parts and operations typify some of the ratiocinative process of understanding psychological phenomena (Wright & Bechtel, 2006). But does that entail that detailed attention to actual scientific practice will portend the sort of tell-all reductionist tale that some might wish to see? I see reason for caution. The case study briefly reviewed here suggests that Bickle's wager isn't a safe bet; the doctrine of psychoneural reductionism is best off purging commitments to *EE*, or at least leaving it at the laboratory door.

In successfully carrying out a reductive explanation of aspects of reward function, researchers often conscientiously work toward a unified, integrated story that bridges the levels inherent in the phrase "neuropsychobiological profile" very seriously (Gallistel, 1994; Ikemoto & Panksepp, 1999; Berridge & Robinson, 2003; McClure et al., 2003; Schultz, 2004).[9] Refinements in reductively understanding the target phenomenon obtain partly because many of the fragmented constructs are characterized across various levels of research, not just

successively lower ones. And the explanation of those constructs characteristically follows a deeply entrenched pattern of terminological and conceptual miscegenation across various levels. So, while psychological explanations of reward function have often played a crucial heuristic role in enabling increasingly accurate neurobiological explanations, the conjunction of that heuristic role and the increased sophistication of lower-level research does not necessarily entail – at least not in this paragon case – that higher-level psychological explanations of paradigmatic functions are generally rendered "impotent" or are left to "disappear" (cf. Bickle, 2003, p. 111). Often, those making Bickle's wager are quick to dispute the charge that they've misunderstood the scientific details. But I don't think that's the problem at all. Psychoneural reductionists understand those details perfectly well, and are certainly right to hold that more attention to them is a good thing; the problem is that selective attention isn't.

## Acknowledgments

This research was supported by a grant from the William J. Fulbright foundation. Conversations about it with William Bechtel, John Bickle, Anthony Landreth, Ioan Muntean, Ken Sufka, Dingmar van Eck, and Iris van Rooij have been extremely rewarding.

## Notes

1 At its most general, "psychoneural reductive explanation" designates a communicative practice that involves one group of cognizers employing various (semiological) conventions to prompt another to produce narrowly interpretable conceptualizations about a psychological-level *explanandum* in terms of a neurobiological-level *explanans*. Typically, this involves both a description of the manner in which higher-level research has initially characterized the occurrence, causal powers, or dynamics of the target phenomenon $\Phi$, and the precise spatial, temporal, and organizational properties, relations, and operations of ontological posits as determined by lower-level research.

2 Feest (2003) argues that, while irreducibility may entail the autonomy/independence of psychological explanation, the converse does not follow; and so reductionists who eschew autonomy because it leads to irreducibility are equivocating between intertheoretic and explanatory autonomy (or, as she puts it, between "explanatory autonomy" and "investigative autonomy"). The argument is plausible, despite the fact that Feest proposes that reductionists are right to claim that psychological explanation fails to be explanatorily autonomous, but wrong to claim that they thereby fail to be intertheoretically autonomous (as the quotation shows, this simply gets the reductionist claim backwards).



3 As always, Bickle has been a forerunner in streamlining the doctrine of psychoneural reductionism – most recently, by turning away from post-classical accounts such as his structuralist New Wave program in order to focus more explicitly on the day-to-day research emerging from cellular/molecular neuroscience. Though he maintains that the formal rendering of the intertheoretic reduction relationship $\rho$ is correct so far as it goes, Bickle's turn toward reductive explanations in cellular/molecular neuroscience is motivated by the recognition that large-scale theories aren't being constructed in actual scientific practice – at least not in the grand sense that was being developed in the New Wave program (personal communication).

4 Note that "extinction" is used as a technical term peculiar to the reductive explanatory patterns inherent in actual scientific practice, and so is *not* synonymous with either "elimination" or "replacement." The latter terms are best reserved for, e.g., intertheoretic contexts where a set of inadequate explanations are unable to be formally reconstructed as part of the set of models and laws comprising some reduced theory $T_R$.

5 Interestingly, the reductive explanation of incentive salience attribution also provides a basis for helping explain reward function gone awry. Di Chiara (1998, 1999) proposed the chronic stimulation of DA neurons in the NAc shell abnormally strengthens stimulus-drug associations of reward-related learning, which artificially confers excessive incentive salience to stimuli predictively associated with some drugs of abuse (e.g., cocaine). This conjecture is buttressed by hypotheses about the dysphoria and anhedonic experiences common in disorders such as depression and schizophrenia. Kapur (2004) proposes that some aspects of psychotic symptoms, for instance, can also be explained by aberrant spiking of DA neurons independent of cue and context; the context-inappropriate release of DA erroneously triggers larger patterns of neurocomputational activation that bring about the confused feelings of altered incentive salience. Over time, the incessant dopaminergic malfunction ends up hijacking normal processes of salience attribution, prediction, and learning. In sum: "Dysregulation of the dopamine system provides the fuel for the creation of the delusion, whereas the patient's personal and cultural

history gives it the precise form" (Kapur, 2004, p. 403). The aberrant salience of psychotic delusions can be attenuated by antipsychotics, whose mechanisms of action include globally damping mesocorticolimbic DA operations (but, unfortunately, thereby downgrades subjects' normal reward function).

6 Since interlevel causation is not necessarily being invoked, *a fortiori* no story about downward causation is required.

7 Some reductionists might be inclined to reply by gesturing at the telling imbalance between higher- and lower-level research (e.g., Bickle, 2003, p. 4). But this would obviously be a red herring: whether neurobiological research on reward function radically "outproduces" its higher-level counterpart is certainly telling, but largely irrelevant to the whether *EE* is true.

8 Motivation constructs are interesting for the case of reward function because they're implicated in a wide range of phenomena – from the intermediate processes involved in incentive salience attribution, to the construction of abstract goals, social ambitions, personal achievement, moral ends, etc. Widening the explanatory scope of reward function research in this way yields yet another reason why contributions from a range of higher-level disciplines are not pointless or silly.

9 Barendregt and van Rappard (2004) welcome the idea of bringing the nature of reductive explanation back into vogue, arguing that the doctrine of psychoneural reductionism has been mired in confusion precisely because of the overemphasis on theory reduction. Instead, they propose that the doctrine is best understood as a general methodological stance which facilitates interlevel cooperation and bridges theories at different levels without ontological concerns. On one hand, what they characterize as "reductive explanation" seems to accurately characterize much of actual scientific practice. On the other hand, what they call "reductive explanation" is actually mechanistic, not reductive, explanation (for explication of the difference between these two types, see Brandon, 1984; Craver, 2001; Wright & Bechtel, 2006). Consequently, if psychoneural reductionist asserting *EE* thereby deny the autonomy/independence of higher-level psychological explanation, this sort of conflation cannot but help fail to make sense of it.

## References


Achinstein, P. (1984). The pragmatic character of explanation. *Philosophy of Science, 2*, S275–S292.

Barendregt, M., & van Rappard, J. F. H. (2004). Reductionism revisited: On the role of reduction in psychology. *Theory and Psychology, 14*, 453–474.





Berridge, K. C. (2000). Measuring hedonic impact in animals and infants: Microstructure of affective taste reactivity patterns. *Neuroscience and Biobehavioral Reviews, 24*, 173–198.

Berridge, K. C. (2003). Pleasures of the brain. *Brain and Cognition, 52*, 106–128.

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology and Behavior, 81*, 179–209.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews, 28*, 309–369.

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences, 26*, 507–513.

Bickle, J. (1998). *Psychoneural Reduction: The New Wave.* Cambridge, MA: MIT Press.

Bickle, J. (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Account.* Dordrecht: Kluwer.

Bickle, J. (2006). Reducing mind to molecular pathways: Explicating the reductionism implicit in "molecular and cellular cognition." *Synthese.*

Brandon, R. N. (1984). Grene on mechanism and reductionism: More than just a side issue. *Philosophy of Science, 2*, S345–S353.

Casassus, G., Blanchet, C., & Mulle, C. (2005). Short-term regulation of information processing at the corticoaccumbens synapse. *Journal of Neuroscience, 25*, 11504–11512.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind–Brain.* Cambridge, MA: MIT Press.

Cohen, J. D., Braver, T. S., & Brown, J. W. (2002). Computational perspectives on dopamine function in prefrontal cortex. *Current Opinion in Neurobiology, 12*, 223–229.

Craver, C. F. (2001). Role functions, mechanism, and hierarchy. *Philosophy of Science, 68*, 53–74.

Craver, C. F. (2003). The making of a memory mechanism. *Journal of the History of Biology, 36*, 153–195.

Di Chiara, G. (1998). A motivational learning hypothesis of the role of dopamine in compulsive drug use. *Journal of Psychopharmacology, 12*, 54–67.

Di Chiara, G. (1999). Drug addiction as dopamine-dependent associative learning disorder. *European Journal of Pharmacology, 375*, 13–30.

Di Ciano, P., Blaha, C. D., & Phillips, A. G. (1998). Conditioned changes in dopamine oxidation currents in the nucleus accumbens of rats by stimuli paired with self-administration or yoked-administration of d-amphetamine. *European Journal of Neuroscience, 10*, 1121–1127.

Feest, U. (2003). Functional analysis and the autonomy of psychology. *Philosophy of Science, 70*, 937–948.

Fiorino, D. F., Coury, A., Fibiger, H. C., & Phillips, A. G. (1993). Electrical stimulation of reward sites in the ventral tegmental area increases dopamine transmission in the nucleus accumbens of the rat. *Behavioral Brain Research, 55*, 131–141.

Fonseca, J. (2004). On Bickle's failure to give a formal account of location in the New Wave reductionist spectrum. *Disputatio, 1*, 50–60.

Freeman, A. S., Meltzer, L. T., & Bunney, B. S. (1985). Firing properties of substantia nigra dopaminergic neurons in freely moving rats. *Life Sciences, 36*, 1983–1994.

Gallistel, C. R. (1994). Foraging for brain stimulation: Toward a neurobiology of computation. *Cognition, 50*, 151–170.

Germans, M. K., & Kring, A. M. (2000). Hedonic deficit in anhedonia: Support for the role of approach motivation. *Personality and Individual Differences, 28*, 659–672.

Heimer, L., Zahm, D. S., Churchill, L., Kalivas, P. W., & Wohltmann, C. (1991). Specificity in the projection patterns of accumbal core and shell in the rat. *Neuroscience, 41*, 89–125.

Hjelmstad, G. O. (2004). Dopamine excites nucleus accumbens neurons through the differential modulation of glutamate and GABA release. *Journal of Neuroscience, 24*, 8621–8628.

Horvitz, J. C., Stewart, T., & Jacobs, B. L. (1997). Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. *Brain Research, 759*, 251–258.

Ikemoto, S., & Panksepp, J. (1996). Dissociations between appetitive and consummatory responses by pharmacological manipulations of reward-relevant brain regions. *Behavioral Neuroscience, 110*, 331–345.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. *Brain Research Reviews, 31*, 6–41.

Kapur, S. (2004). How antipsychotics become anti-"psychotic" – from dopamine to salience to psychosis. *Trends in Pharmacological Sciences, 25*, 402–406.

Kiyatkin, E. A. (1995). Functional significance of mesolimbic dopamine. *Neuroscience and Biobehavioral Reviews, 19*, 573–598.

McClure, S. M., Daw, N. D., & Montague, P. R. (2003). A computational substrate for incentive salience. *Trends in Neurosciences, 26*, 423–428.

Neill, D. B., Fenton, H., & Justice, J. B. (2002). Increase in accumbal dopaminergic transmission correlates with response cost not reward of hypothalamic stimulation. *Behavioral Brain Research, 137*, 129–138.

Nicola, S. M., Taha, S. A., Kim, S. W., & Fields, H. L. (2005). Nucleus accumbens dopamine release is necessary and sufficient to promote the behavioral response to reward-predictive cues. *Neuroscience, 135*, 1025–1033.



Pfaus, J. G., Damsma, G., Wenkstern, D., & Fibiger, H. C. (1995). Sexual activity increases dopamine transmission in the nucleus accumbens and striatum of female rats. *Brain Research, 693*, 21–30.

Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioral mechanisms of reward and motivation. *Current Opinion in Neurobiology, 6*, 228–236.

Robinson, T. E., & Berridge, K. C. (2000). The psychology and neurobiology of addiction: an incentive-sensitization view. *Addiction, 95*, S91–S117.

Robinson, T. E., & Berridge, K. C. (2004). Incentive-sensitization and drug "wanting." *Psychopharmacology, 171*, 352–353.

Salamone, J. D., & Correa, M. (2002). Motivational views of reinforcement: Implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behavioral Brain Research, 137*, 3–25.

Salamone, J. D., Correa, M., Mingote, S. M., & Weber, S. M. (2005). Beyond the reward hypothesis: Alternative functions of nucleus accumbens dopamine. *Current Opinion in Pharmacology, 5*, 34–41.

Schouten, M. K. D., & Looren de Jong, H. (1999). Reduction, elimination, and levels: The case of the LTP-learning link. *Philosophical Psychology, 12*, 237–262.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1–27.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron, 36*, 241–263.

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioral ecology. *Current Opinion in Neurobiology, 14*, 139–147.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*, 1593–1599.

Schweizer, P. (2001). Realization, reduction, and psychological autonomy. *Synthese, 126*, 383–405.

Spanagel, R., & Weiss, F. (1999). The dopamine hypothesis of reward: Past and current status. *Trends in Neuroscience, 22*, 521–527.

Taha, S. A., & Fields, H. L. (2005). Encoding of palatability and appetitive behaviors by distinct neuronal populations in the nucleus accumbens. *Journal of Neuroscience, 25*, 1193–1202.

Ungless, M. A., Magill, P. J., & Bolam, J. P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science, 303*, 2040–2042.

Wise, R. A. (1998). Drug-activation of brain reward pathways. *Drug and Alcohol Dependence, 51*, 13–22.

Wise, R. A. (2002). Brain reward circuitry: Insights from unsensed incentives. *Neuron, 36*, 229–240.

Wise, R. A. (2004). Dopamine, learning, and motivation. *Nature Reviews Neuroscience, 5*, 483–494.

Wright, C. D. (2000). Eliminativist undercurrents in the new wave model of psychoneural reduction. *Journal of Mind and Behavior*, *21*, 413–436.
Wright, C. D. (2002). Animal models of depression in neuropsychopharmacology qua Feyerabendian philosophy of science. In S. P. Shohov (Ed.), *Advances in Psychology Research* (Vol. 13, pp. 129–148). New York: Nova-Science Publishers.
Wright, C. D., & Bechtel, W. (2006). Mechanisms and psychological explanation. In P. Thagard (Ed.), *Handbook of Philosophy of Psychology and Cognitive Science* (pp. 31–79). New York: Elsevier Science.

# 12

# WHO SAYS YOU CAN'T DO A MOLECULAR BIOLOGY OF CONSCIOUSNESS?

*John Bickle*

## 1. "Molecular Consciousness Studies"?

The answer to this chapter's title is: Pretty much everybody now working in consciousness studies. This includes philosophers whose attacks stem from global denials about any physical science accounting for some aspects of conscious experience. Colin McGinn is a clear example: "We are cut off by our very cognitive constitution from achieving a conception of that natural property of the brain (or of consciousness) that accounts for the psychophysical link" (1989, p. 350) – molecular biological properties among these. But it also includes less skeptical philosophers and scientists who seek explanations of consciousness at higher levels of biological organization, such as information processing (e.g., Baars, 1988), neural networks (e.g., Paul Churchland, 1995; Patricia Churchland, 2002), and clinical neuropsychology and neurology (e.g., Ramachandran & Blakeslee, 1998; Damasio, 1999).

Even the handful of basic neuroscientists who concern themselves with consciousness say little to nothing in print about potential molecular mechanisms. Francis Crick and Christof Koch's now-classic essay (1990) is a good example. They are not mysterians: "We suggest that time is now ripe for an attack on the neural basis of consciousness" (1990, p. 264). Nor are they cognitivists: "Arguments at the cognitive level are undoubtedly important but we doubt whether they will, by themselves, ever be sufficiently compelling to explain consciousness in