On the non-elimination of mental states by adopting a ruthless-reductive stance

In several places, John Bickle claims that current neuroscientific practice provides actual cellular/molecular reductions of certain mental states. He gives the case study of ‘memory consolidation switch’ as an example where recent findings suggest that this mental state/process can be reduced to the molecular ‘cAMP, PKA, CREB Pathway’. Taking this example, Bickle ‘waves the eliminativist flag’ by claiming that psychological explanations lose their pertinence (or, as he says, ‘became otiose’) once a cellular/molecular explanation replaces them. On this paper I’ll try to show that, even if a reductive explanation of ‘memory consolidation switch’ is disposable, we cannot eschew reductively its causal-functional integrity, i.e.: the explanatory/causal context that defines the mental concept/process ‘memory consolidation switch’ in the first place.

Intentional/Cognitive Models in Current Neuroscientific Practice

I will adopt here a model/semantic approach to neuroscientific practice (specifically, on the present context, that part of neuroscientific practice dealing with the explanation of behaviors, hence, I will call this kind of activity as Cognitive Neuroscience – or CN for short). This semantic view has some connections to Ronald Giere’s approach to physics and has noting to do with the structuralist framework sometimes endorsed by Bickle himself.

Models of CN are intentional objects, cognitive representations used by scientists in several ways (they can be represented verbally graphically or in another way). More precisely, CN models are intended abstractions of neuroscientists empirical results/findings. Frequently, these intentional models are abstractions of animal experimental models. For instance, protocol experiments of fear condition on mice are interpreted/intentioned as a model of fear conditioning on mammals (at least). Different animal models can share the same intended model, eg; fear conditioning in rabbits have the same intended model as the ones on mice. The different animal models are clustered together under the same abstract model by a relation of similarity (the relevant similarity is established by the intended perspective adopted – on this case, the understanding of fear conditioning in mammals). The general empirical and extensional domain of all these models is the set of terrestrial animals.
Another important aspect concerning this kind of models is the fact that different models can have different ‘grains’ of explanation. This concept of ‘grain’ has been introduced and developed by William Bechtel and Jennifer Mundale in their challenge to the Multiple Realization argument. Basically, what they correctly stress is that neuroscientists seek explanations of cognitive functions in different degrees of complexity and intended target scope (more ‘fine or coarse grained’ explanations). Bechtel and Mundale upshot is that, in discussing neural realization of mental concepts, we should have in consideration this differences in grain both from the mental side but in the neuronal side as well and find the correct level of grain in any mental-neural identity or correlation. Translated to the current context, it means that CN models can vary in their ‘grain’ and, therefore, in their intended scope, eg; a model of ‘contextual fear condition’ is more fine grained and with a more restricted scope than a model of ‘fear conditioning’ simpliciter. Hence, a consequence of this view is that models can ‘overlap’ in the sense that a certain model can ‘incorporate’ a more coarse grained model (this happens in the former example of ‘contextual fear conditioning’ and ‘fear conditioning’). Here, again, it parallels Giere’s approach to models in classical mechanics. For instance, a model of Damped Pendulum is simpler (less fine grained) than a model of Damped Driven Pendulum and the latter ‘includes’ the former.

Now is time to be more precise about models structure in order to shed some light on what has been said so far (but only by ‘scratching the surface’ of this framework). CN models have two basic ‘components’: a measured behavior $B$ and a description of a neural/componential mechanism $M$ causally explaining $B$ (for obvious reasons of lack of time and space I won’t elaborate here on the interesting questions concerning the epistemological and ontological status of mechanistic explanations). The explanatory target of these models (as a compound ‘entity’ of this two components) is a certain mental state/process $P$ previously defined as assuming a certain role in explaining $B$ (in some respects this view about $P$ is akin to a functionalist conception minus its metaphysical consequences regarding mental states). So, in a certain sense we can evoke ‘identity’ holding between $P$ and $M$, not in a metaphysical sense but as a pragmatic role endorsed within scientific practice or, what Bechtel and McCauley call ‘Heuristic Identity Theory’ (this also stresses that these models favors - and are part of - a dynamics of co-evolution between mental states/psychology and neuroscientific explanation). There are CN models of ‘Fear Conditioning’, ‘Spatial Memory’, ‘Memory Consolidation’ and so on. Neuroscientists try to discover what are the neural causal
mechanisms ‘realizing’ these mental states/functions by empirically search for the mechanisms $M$ that assume the explanatory role of $B$ that defines $P$.

I should now introduce an important point concerning the nature of mental states $(P)$ in the present context in order to avoid confusions and misunderstandings. I am assuming a realist stance towards mental states. By ‘realism’ I mean that neuroscientists conceive mental states as (supposedly) ‘realized’ by certain neural/componential mechanisms that they try to discover. Not only scientists constantly assume this but there are important methodological and pragmatic reasons to act this way. The main reason behind this assumption is ‘explanatory relevance’. Let me try to clarify this point by providing an actual example. Let consider again the case of Fear Conditioning. In a series of places the well known neuroscientist Joseph LeDoux (the ‘father’ of modern studies of fear conditioning, especially in mammals) says that fear conditioning is observable in a wide range of animals, from the Drosophila to humans. But this is an error! What LeDoux should say is that avoidance behavior is observable in all those organisms. Fear Conditioning explains (or is supposed to) avoidance behavior. But doesn’t the Drosophila exhibits avoidance behavior (as the famous experiments of Tim Tully supposedly show)? Yes, but first of all, what Tully (and other researchers) were intending with the Drosophila’s ‘avoidance task’ was not ‘fear conditioning’ (or even avoidance conditioning for that matter) but ‘memory’ (short term, long term and consolidation) and found important molecular and genetic mechanisms shared by other animals including mammals. The ‘avoidance protocol’ was chosen because of its pragmatic advantages, there could have been other experimental settings (measuring different behaviors) to arrive at the very same conclusions. If we search for the ‘fear mechanism’ that explains the Drosophila avoidance behavior we have to include the memory mechanism into a complex and ultra-detailed description including how specific sensory inputs and specific motor effectors that produce the behavior. All this (and this is crucial for the present context) would be ‘Drosophila specific’, i.e., that explanation could not be extrapolated and generalized outside the Drosophila ‘domain’, in sum, it is too ‘fine grained’. Here enters the ‘explanatory relevance’ constraint: scientists expect their explanations to have the wider domain of application possible. The identification of the behavior with the mental state has the ‘balkanization’ consequence we just testified. The ‘realist’ stance towards mental states is not a whim but a methodological golden rule in contemporary neuroscientific practice as LeDoux himself exemplifies on his work on fear conditioning by suggesting a neuronal
mechanism (the amygdaloid structure) as the ‘locus’ of ‘fear conditioning’ at least in mammals. So, from the neuroscientific practice perspective, mental states are, not only ‘real’, but also possess ‘projectability’ (answering, in some way, to a well known metaphysical anxiety felt by Jaegwom Kim).

So, applying the ‘explanatory relevance’ constraint (and a little bit of philosophical simplification): neuroscientists assume a previous definition of a mental state/process $P$ in terms of its explanatory role of a certain behavior $B$ and not by an identification of $P$ with $B$ itself. From this, they search for the neural mechanisms $M$ that equally explain $B$. They do this by constructing experimental settings that exemplify the target behavior $B$ and detect the structure $M$ by several intervention techniques that change $B$ in relevant ways that makes them conclude that $M$ is the structure underlying $P$. The grade of detail of $M$ surly varies as the model is more or less coarse-grained. In any case, the mechanistic explanation of $P$ (and hence of $B$) will always be ‘gappy’ or incomplete. Carl Craver calls this incomplete mechanisms descriptions ‘sketches’, i.e., descriptions of mechanisms deploying black boxes or filler terms in them. There are degrees of ‘sketches’, more or less incomplete. For my purpose here, what is important is that a sketch of a certain mechanism as deployed in a given model can be ‘filled in’ by adopting a more fine-grained model that overlaps the former. This can happen if we deliberately reduce the intended scope of the model, eg from ‘short-term memory’ simpliciter to ‘short-term memory in mammals’ or by qualifying it as in the case referred above of ‘fear conditioning’ and ‘contextual fear conditioning’.

The question of interest is that, when scientists intend finer-grained overlapping models, the mental state/process being modeled is not the same (‘short-term memory’ is not the same mental state as ‘short-term memory in mammals’).

The functional integrity of ‘memory consolidation switch’

John Bickle purpose is to show how current neuroscience provides reductions at the level of the cellular-molecular of certain mental states/processes. He provides the example of the supposed reductive “‘Memory Consolidation Switch’ (MCS) – ‘cAMP, PKA, CREB, Pathway’” link. Not just that, Bickle argues that this accomplished reduction renders psychological explanations ‘otiose’.

According to the framework just sketched, Bickle seems to claim two different things: (1)- He means that MCS is reduced as $P$ to a neuronal/componential (cellular-
molecular) mechanistic explanation \( M \) and not just to a functional-cognitive/non-componential explanation \( M' \) of MCS, and (2)- He is saying that MCS’s explanatory role of a certain behavior \( B \) becomes otiose once we get the reduction mentioned in (1). Although not always clear, sometimes Bickle seems to endorse (2) as a corollary of (1). Let’s take a closer look at both (1) and (2).

In relation to (1) Bickle is ‘crystal clear’; there was a cognitive/functional explanation of MCS that became reductively suppressed by a more detailed and neuronal (cellular-molecular) successor. The previous explanation of MCS was based on psychological experiments and can be crudely summarized as follows: in order for a certain ‘memory trace’ to consolidate from short-memory to long-term memory it has to be repeated \( n \) times during the relevant ‘consolidation period’ without any ‘retrograde interference’. This would be the functional and non-componential explanation of MCS. Bickle characterizes this sort of explanation (within a structuralist model-theoretic framework of inter-theoretic reduction) as follows:

Models of a psychological theory of memory […] posit an entity/process, the consolidation switch. But they characterize this posit only in terms of the time course and amount of repetition needed to convert a type of memory item from short-term memory to long-term memory and the behavioral efficacy of different types of retrograde interference. In other words, psychology characterizes this entity/process in purely functional fashion (Bickle, 2003, 99)

He goes on by showing how this purely functional explanation gets reduced by current cellular-molecular neuroscience, specially by the recent findings on the underlying cellular and molecular mechanisms of ‘Long Term Potentiation’ (LTP), in particular the mechanisms of the extension of Early Phase-LTP (E-LTP) into Long Phase-LTP (L-LTP):

The consolidation switch empirical base set in models of psychology got mapped to sequences and combinations of empirical base sets and fundamental relations in reduction-related models and intended empirical applications of molecular neuroscience, in particular to those involved in the transition of E-LTP into L-LTP and the maintenance of L-LTP. The empirical base sets of the latter include intracellular and neural transmission molecules: adenyl cyclase, cAMP, PKA, CREB enhancers and repressors, DNA, RNA polymerases, ubiquitin hydrolase, CCAAT enhancer biding protein, glutamate, dendritic spine cytoskeleton components, AMPA receptors, NMDA receptors, and so on.
So, basically, Bickle claims that the functional/information-processing explanation of MCS is replaced by a better neural (cellular-molecular) correlate. This process, Bickle claims, is a ‘reduction’, in particular a reduction of a certain previous functional explanation $M'$ of $P$ (MCS in this case) for a new componental/neuronal one $M$. This is an important advance in terms of constructing a CN model for MCS, since these models are neuroscientific, i.e., seek componental-neuroscientific mechanistic explanations $M$ of mental states $P$ and not purely functional ones. This is basically what (1) states.

So, what about (2)? It should be stressed that Bickle himself is not completely clear concerning this point. Although sometimes, when he talks about ‘replacing’ psychological explanation he makes that clearly in the context just presented of a reduction/replacement of the functional explanation of MCS by a neuroscientific one, in other occasions he seems to consider MCS integrity as a mental state also to be the target of the reductionist ‘sweeping’. That is particularly clear when he discusses the topic of Mental Causation within this particular context. Concerning mental causation, he says that:

I contend that when we fix our gaze on aspects of scientific practice in this actual recent example, we see that psychological explanations lose their initial status as causally-mechanistically explanatory vis-à-vis as accomplished (and not just anticipated) cellular/molecular explanation. All attempts by philosophers to “save” mental causation presuppose that psychological explanations remain (causally) explanatory (2003, 110, emphasis in the original)

Where ‘mental causation’ does enters the picture? (I’m not interested here on metaphysical/philosophical conceptions and discussions concerning the nature of causation but in a more intuitive and prosaic notion as many time adopted by scientists which is the context here in discussion) Recall that in any CN model the target mental state $P$ is defined/identified by its causal/explanatory role of the behavior $B$. In the particular case of MCS we can, tentatively, establish an explanatory relation as something like: ‘MCS’ explains/causes ‘Long time preservation of former acquired behaviors’. Bickle’s idea concerning this point seems to be that, by achieving a ‘low-level’ reduction of MCS, its overall functional/causal integrity would dissipate in the wider cellular-molecular explanation. So, (2) would derivate from (1). After describing
the cellular and molecular components of L-LTP maintenance in the struturalist/reductionist framework (quoted above), he adds the following:

Out of these components, the fundamental relations of cellular/molecular neuroscience build membranes, plastic neurons, neuronal circuitries from sensory to motor effectors, neuromuscular junctions, and the specific connections with the muscles and skeletal systems (themselves build up by fundamental relations out of their molecular elements). (2003, 99)

In a more recent paper, Bickle’s strategy for deriving (2) from (1) is even clearer:

[T]he cellular or molecular events in specific neurons into which experimenters have intervened, in conjunction with the neuronal circuits in which the affected neurons are embedded, leading ultimately to the neuromuscular junctions bridging nervous and muscle tissue, directly explain the behavioral data. (2006, emphasis in the original)

So, Bickle thinks that a ‘low-level’ explanation of MCS in terms of cellular/molecular mechanisms $M$ «in conjunction with» an embedment of $M$ within a wider and detailed mechanism $M''$ of MCS and a description of how sensory inputs and motor outputs are carried out at a very fine grained level provides a reduction of the input-output functional relations characterizing MCS

But Bickle is incurring in a fundamental error. We are dealing with a CN Model, a very ‘coarse-grained’ model of MCS. Bickle himself acknowledges this (even if implicitly) by noting that MCS is, not only a property of hippocampal LTP, but also an underlying phenomenon of many types of memory (declarative, non-declarative). Because of this, he reframes the fundamental mechanism commonly explaining all cases of MCS though the animal kingdom: «There is a “physical-chemical state” the cAMP-PKA-CREB molecular biological pathway, that uniquely realizes memory consolidation across biological classes, from insects to gastropods to mammals» (2003, 148). Of course, the ‘cAMP-PKA-CREB pathway’ is an oversimplification of a more complex mechanism, although a very gappy one (a very incomplete mechanism sketch) full of ‘black boxes’, not by ignorance but because it is an abstraction of several particular more complex mechanisms. Bickle asserts that this pathway is common to all animals displaying MCS (aided on this conclusion by a phylogenetic conservation argument) and this is true despite «vast differences in brain size, organization, site of
principal effect (presynaptic or postsynaptic)» (2003, 148). He appeals to the ‘minimum common denominator’ of MCS in order to achieve the wider explanatory scope possible. Thus, we got a sort of ‘partial reduction’ of MCS but, nonetheless, suitable for the ‘coarse-grained’ model in question that, by itself, provides a frame to be filled in by more fine grained overlapping models (qualifying in more particular ways the MCS state). So, as Bickle recognizes, for the sake of explanatory relevance of a model of MCS we cannot embed \( M \) within \( M' \). Furthermore, and for the same reasons, Bickle has to restrain his reductive ambitions concerning MCS input-output profile. We cannot provide a detailed and fine-grained description of the sensory inputs and motor outputs in the model if we want to keep it empirically correct, i.e., as intending to cover all the elements of the ‘general domain’ – terrestrial animals - satisfying the open sentence ‘\( x \) Displays MCS’. There is too much variation in sensory and motor pathways and implementational tissues to even think about that possibility. A direct explanation of the behavioral data, as envisioned by Bickle in the quoted passage, would only be Token-explanatory, and this is a limitation in scope that Bickle himself explicitly avoids (Bickle can’t have it both ways).

So, although we can provide a ‘minimal’ cellular/molecular mechanism of MCS in the model, MCS functional and causal definition prevails as before (“‘MCS’ explains/causes ‘Long time preservation of former acquired behaviors’”) since MCS’s functional profile is Multiple Realized. Therefore, contrarily to Bickle’s ambitions (2) do not follows from (1) and this conclusion undercuts the eliminativist warrants that he seems to endorse regarding MCS.