The bridge between neuroscience and cognition must be tethered at both ends

Oren Griffiths, Mike E. Le Pelley and Robyn Langdon

Griffiths, O., Langdon, R., & Le Pelley, M. E. (2015). The bridge between neuroscience and cognition must be tethered at both ends. *Cognitive Neuropsychiatry, 20*, 106-10, doi: 10.1080/13546805.2014.993464.

COMMENTARY: The bridge between neuroscience and cognition must be tethered at both ends.

Oren Griffiths[1], Mike E. Le Pelley[1,2] and Robyn Langdon[2]

1. School of Psychology, UNSW Australia, Sydney, Australia

2. Australian Research Council (ARC) Centre for Cognition and its Disorders, Sydney, Australia

Corresponding author:

Dr Oren Griffiths

School of Psychology,

UNSW Australia,

Sydney, NSW, 2052, Australia

Ph: +61 2 9385 3041

Fax: +61 2 9385 3641

Email: oren.griffiths@unsw.edu.au

Running Head: INTERPRETING NEURAL CORRELATES

Keywords: Schiozophrenia, fMRI, delusions, associative learning, prediction error.

Word count: 1000.

We have enjoyed this dialogue with Corlett and colleagues (Corlett & Fletcher, in press; Griffiths, Langdon, Le Pelley, & Coltheart, 2014), which we hope has helped clarify issues regarding prediction error theories of psychosis, and the manner in which empirical data have been interpreted to support those theories. On reviewing the debate, we find we are in agreement on many issues discussed, and disagree on only a small, but important, minority.

The first point of agreement is an interest in prediction error theories, and also the suggestion that functional brain systems may not be neatly compartmentalized, but instead interact to yield mental processes. We also remain invested in the concept of a hierarchy relating perception to cognition. A point of disagreement, though, relates to Corlett and Fletcher's argument that *any* division drawn between perceptual and inferential processes is unhelpful. We see this conclusion as a rather extreme interpretation of the observation that perceptual and inferential levels often interact.

We also agree that bridging cognition with neural science is a difficult endeavour. It requires mapping latent cognitive constructs (e.g., beliefs) with activity in an organ that is difficult to measure online. A further complication arises when using this methodology to examine psychotic phenomena. This area is fraught with definitional problems (what exactly defines a delusional belief? David, 1999), disagreement about the causative nature of the problem (is the abnormality in perception, inference, or both?), and is plagued by confounds, including the effects of disease progression and treatment on cognition (Meltzer & McGurk, 1999).

Like Corlett and colleagues, we believe empirical techniques derived from associative learning theory (ALT) provide a useful set of tools to aid this research. The utility ALT offers is that it clearly defines latent variables and their interactions. For example, prediction error (a latent variable) is the difference between a cue's associative strength (also a latent

variable) and the intensity of the outcome (an observable variable). Most importantly, ALT anchors these mental processes or interactions in well-established behavioural phenomena, such as the blocking task. It thereby grounds hidden mental processes (e.g., belief updating) in observable behavioural terms, allowing one to draw inferences of association between mental processes (assumed on the basis of a closely-related measured behaviour) and concurrent neural activity (measured independently).

Many cognitive neuroscience researchers, including Corlett and colleagues, are now taking up the challenge of studying psychotic symptoms using ALT procedures and constructs. As Corlett et al. (2004) note, a retrospective revaluation procedure (when properly conducted) is an elegant ALT procedure that controls many of the variables that are important to exclude in imaging studies (frequency of exposure, the stimuli present onscreen at test, etc). The key problem is that, in the target studies we reviewed that examined the effects of ketamine and psychosis, Corlett et al. (2006, 2007) did not analyse their neural data in a way that adequately measures a retrospective revaluation effect (see Griffiths et al., 2014, for details). So how can we confidently link the neural activity identified by Corlett et al. (2006, 2007) with a mental process (belief revision) indexed by an effect (retrospective revaluation) that was not observed?

Corlett and Fletcher, in their response, state that their interpretation of these data used reverse inference. Specifically, they analyzed the neural data using an inherently ambiguous contrast which controlled for baseline activity. This meant that the two contrasted cues differed, not only with regard to their potentials for belief revision, but also in many other ways. Consequently, the one region (rDLPFC) whose contrast was differentially activated across groups may indicate differential prediction error signaling, or it may indicate some other group difference (working memory processes, for example). To deduce which mental function the observed differential neural activity in rDLPFC was associated with, Corlett et al

then refer to earlier research which claimed to have localized belief-updating activity to the rDLPFC. On these grounds they inferred that their data provided a neural index of belief-updating. Corlett et al argue that, although this kind of reverse inference has limitations, it is sometimes appropriate.

We agree that reverse inferences can be informative and that convergence across studies can advance understanding. We depart from Corlett and Fletcher in remaining concerned that, in this instance, reverse inference was unnecessary. Specifically, the retrospective revaluation procedure used by Corlett et al. (2006, 2007) in their studies of psychosis and ketamine has a contrast built into the behavioural evidence (the comparison of backward blocking and unovershadowing conditions) which would not require a reverse inference to interpret, since it controls elegantly for all factors other than belief revision. But in each of their studies, the authors ignored this contrast in favour of analyzing one that was less well-controlled and required a potentially ambiguous reverse inference to interpret. So, we remain concerned that their interpretation (which claims that their study indexed belief revision, that it is localized to a particular cortical region, and that it is disrupted by the presence of schizophrenia or ketamine), goes beyond what their data warrant.

One may wonder whether this matter is simply a difference of opinion regarding how to interpret a small set of studies. But the implication is wider because it highlights an issue integral to the interpretation of neuroimaging data in cognitive terms. A central benefit of ALT is that it defines mental functions by reference to measurable, behavioural effects. If one fails to consider the most relevant behavioural effects, then in the absence of other independent evidence (i.e., not based on neural activity) that a mental function of particular interest is being indexed, one loses the ability to draw strong inferences about the relationship between that mental function and its underlying neural substrates. We agree with Corlett and Fletcher's observation that "it is not clear exactly how we can make bridges between

cognitive and neural science". We also acknowledge the need for an eclectic methodological tool-bag to pursue this goal. However, we remain convinced that in order to bridge brain (neural activity) with mind (cognitive functions), one should aim to anchor that bridge in observable phenomena at both ends.

References

Corlett, P. R. & Fletcher, P. C. (in press) Delusions and prediction error: Clarifying the roles of behavioral and brain responses. *Cognitive Neuropsychiatry.*

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A. E., et al. (2004). Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron, 44*, 877–888.

Corlett, P. R., Honey, G. D., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Absalom, A. R., et al. (2006). Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine. *Archives of General Psychiatry, 63*, 611–621.

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., et al. (2007). Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. *Brain, 130*, 2387–2400.

David, A. S. (1999) On the impossibility of defining delusions. *Philosophy, Psychiatry, & Psychology*, 6, 17-20

Griffiths, O., Langdon, R., Le Pelley, M.E. & Coltheart, M. (2014) Delusions and prediction error: Re-examining the evidence for disrupted error signaling in delusional people. *Cognitive Neuropsychiatry*, 19, 439-467.

Meltzer, H. Y. & McGurk, S. R. (1999) The effects of clozapine, risperidone, and olanzapine on cognitive function in schizophrenia. *Schizophrenia Bulletin*, 25, 233-255.

Acknowledgments

This work was supported by grants from the National Health and Medical Research Council [grant number APP1069487] and the Australian Research Council [grant number FT100100260].