U-turns in the brain

When making decisions, new information sometimes calls for a change of mind. New results indicate that regions of the prefrontal cortex play distinct roles in evaluating new evidence in light of a previous choice and translating the result of this evaluation process into an explicit report of one's subjective confidence.

Redmond G. O'Connell and Peter R. Murphy

If you never change your mind, why have one?
—Edward De Bono

Throughout the course of his campaign for the US presidency in 2004, John Kerry was plagued by accusations that he was an unreliable "flip-flopper" because he had altered his position on certain policy issues over the course of his political career. The knockout blow of a famous attack ad at the time was that John Kerry "has changed his mind time and time again." In a society that tends to stigmatize indecision, it is easy to overlook the fact that our survival and well-being require that we continually evaluate and revise our decisions as new evidence comes to light; when we step off the sidewalk to cross the road in the belief that the coast is clear, we must have the capacity to quickly reverse that decision if a speeding car suddenly rounds the bend. A new study reported in this issue of *Nature Neuroscience*[1] combines clever experimental design with functional neuroimaging and computational analysis to shed light on how the brain enables us to achieve this indispensable cognitive flexibility.

A long-standing proposal has been that the brain employs specialized 'metacognitive' mechanisms that monitor and regulate the primary neural processes through which decisions are formed[2]. Indeed, functional neuroimaging studies have consistently highlighted a set of primarily frontal brain areas whose activation patterns are more closely tied to error detection and fluctuations in choice confidence than to the associated decisions[3–5]. Meanwhile, depending on its location, brain damage or disruption can impair the accuracy of our decisions without necessarily affecting our ability to evaluate them and vice versa[6–8], suggesting that distinct mechanisms are invoked for making and monitoring decisions. Beyond the idea that certain brain areas appear to be specialized for decision monitoring, however, the cognitive operations performed at this processing level have remained mysterious.

Fleming et al. set out to test the hypothesis that one such operation computes the degree to which new information conflicts with an existing belief and thereby furnishes a neural signal that can be used to inform changes of mind. Examining this possibility required very careful study design, the details of which are worth highlighting. Participants were asked to make difficult perceptual judgments regarding the dominant direction of motion (leftward or rightward) in a cloud of moving dots, after which they rated their confidence in their choice on a continuous scale that ranged from 'certainly correct' to 'certainly wrong' (the latter indicating a definite change of mind; Fig. 1). Such elementary perceptual discrimination tasks have been employed extensively by neuroscientists seeking to isolate the core computations the brain implements to make decisions[9]. In a clever twist on this type of task, Fleming and colleagues allowed their participants to briefly view the dot-motion stimulus a second time, after committing to their decision but before reporting their confidence. The dots during this second viewing period always moved in the same dominant direction as those in the first, but could do so more obviously or less obviously. This ensured that participants were exposed to varying levels of post-decisional evidence.

Another important design feature of the study was that participants were rewarded, not according to the accuracy of their motion discriminations, but according to the accuracy of their confidence ratings. Thus, low reported confidence in an incorrect choice yielded as high a payoff as high confidence in a correct one. This approach allowed the authors to precisely quantify the impact of post-decisional evidence on confidence reports and brain activity in a manner that excluded potentially confounding influences of trial difficulty and value.

As predicted by a computational model of both choice and confidence on this task, the participants' confidence reports were highly sensitive to the post-decisional evidence, with stronger motion leading to lower confidence in incorrect choices and higher confidence in correct ones. Fleming et al. then examined functional imaging data from the same participants and found that activation of the posterior medial frontal cortex (pMFC)

specifically scaled with the likelihood that a previous decision was incorrect given the new, post-decisional evidence. This demonstration suggests that a core function of the pMFC may be to signal when newly received information calls for a current belief to be revised, an account that can reconcile a great many findings pertaining to this brain region, including its activation during the detection of errors, during conflict between incompatible responses that are activated simultaneously, and when deciding in foraging contexts to test alternative options in the face of diminishing returns.

These analyses also highlighted a distinct role for the lateral anterior prefrontal cortex (aPFC). Previous work by Fleming and colleagues[10] has demonstrated that, while this region stays quiet when people make decisions without reporting their confidence, it activates strongly when they are required to give explicit confidence reports. Here they build on this finding by showing that the lateral aPFC mediates the relationship between post-decision evidence and subsequent confidence ratings. This suggests that, while pMFC represents the likelihood that a decision is incorrect, lateral aPFC is involved in translating this information into an explicit report.

The work of Fleming et al. provides important insights into how our brains represent decision evidence. Whereas decisions are initially formed by appraising sensory evidence with respect to the choice alternatives (probability of leftward versus rightward motion), the regions of prefrontal cortex highlighted here appear to employ a distinct frame of reference for evidence that arrives after the point of commitment (probability that a decision is correct versus incorrect). It is tempting to infer from this that such metacognitive representations play a direct role in the implementation of changes of mind, which would be consistent with related computational accounts in which pMFC signals provide the input to error detection processes[5]. However, the new results do not rule out alternative possibilities. Complementary computational modeling and neurophysiological research has indicated that decisions are formed in specialized circuits that accumulate the
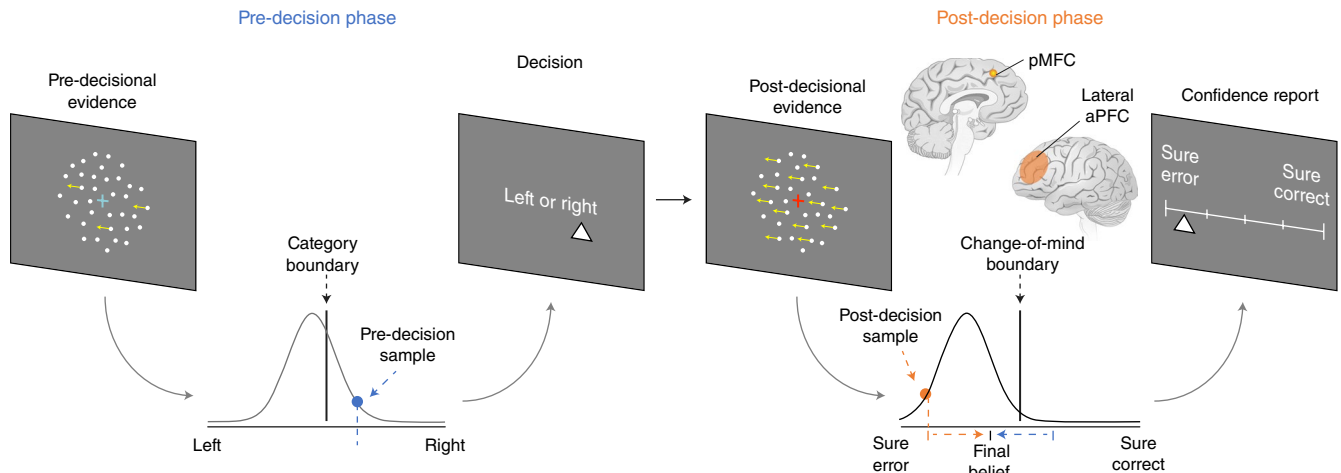
Fig. 1 | Neural mediators of changes of mind. On each trial, participants viewed a cloud of moving dots. Most of the dots moved randomly, but a subset moved coherently either to the left or to the right, and participants were asked to report the dominant direction of coherent motion. In this example, the dominant motion direction presented during the initial pre-decision phase is leftward, but the participant incorrectly perceives it to be rightward. In the post-decision phase, participants were asked to report their confidence in their choice using a sliding scale (sure error to sure correct) but were first allowed to view the dot cloud again. The dominant motion direction was always the same as that presented in the pre-decision phase, but could be more or less obvious (i.e., more or fewer dots moving coherently). In this example, the strong post-decisional evidence causes the participant to realize that their initial decision was likely to be incorrect. Functional MRI analyses revealed that the pMFC represented the degree to which the post-decision evidence conflicted with the original choice, while the lateral aPFC mediated the impact of post-decision evidence on reported confidence. Credit: Marina Corral Spence/Springer Nature

available evidence in favor of each choice alternative and trigger commitment to one of these alternatives once a sufficient quantity of evidence has accrued[9]. If contradictory evidence is encountered following commitment, these same circuits have the capacity to reverse a decision without recourse to higher-level metacognitive representations[11,12]. This being so, what additional benefits might be conferred by re-representing this information in metacognitive circuits?

One possibility is that metacognitive signals have no direct impact on choice behavior at all and are designed primarily to facilitate the communication of our internal mental states to other people. However, numerous functional MRI and electrophysiological studies have reported that medial prefrontal signals predict behavioral adjustments following errors, suggesting that they influence future decisions[13,14]. Moreover, recent human electrophysiology work has demonstrated that these signals emerge even while the initial decision is still being formed[15], indicating that they are also well positioned to influence evidence accumulation processes in real time. Thus, rather than directly driving changes of mind, representations of choice confidence may facilitate them by determining the extent to which we opt to persist in gathering

evidence after choice commitment or move on to other tasks. Key to resolving these questions will be to finely trace the temporal evolution of neural signatures of decision formation and metacognition in parallel and to relate their dynamics to variations in both current and future choice behavior.

In most experimental tasks, decision-making is operationalized as a discrete process that terminates in a commitment to a particular belief, but the reality is that decision-making can be a fluid process, one of continual vacillation in which our beliefs are reinforced, revised, or abandoned as new evidence is encountered. By drawing research on metacognition into the same mechanistic framework that has proven so fruitful in illuminating the neural processes underpinning perceptual decision-making, researchers like Fleming et al. are making strides toward characterizing the computations that are performed at this high level of processing. Of course, there is much work yet to be done before we reach a full understanding of precisely how metacognitive mechanisms influence our decisions, and we will likely have to make some U-turns in our thinking along the way. ❑

Redmond G. O'Connell[1]* and Peter R. Murphy[2]

[1]*Trinity College Institute of Neuroscience and School of Psychology, Trinity College Dublin, Dublin, Ireland. [2]Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.*
*e-mail: reoconne@tcd.ie

References
1. Fleming, S. M., van der Putten, E. J. & Daw, N. D. *Nat. Neurosci.* https://doi.org/10.1038/s41593-018-0104-6 (2018).
2. Nelson, T. O. & Narens, L. Metamemory: a theoretical framework and new findings. in *The Psychology of Learning and Motivation: Advances in Research and Theory* (ed. Bower, G.H.) 125–173 (Academic, New York, 1990).
3. Ridderinkhof, K. R., Ullsperger, M., Crone, E. A. & Nieuwenhuis, S. *Science* **306**, 443–447 (2004).
4. Fleming, S. M. & Dolan, R. J. *Phil. Trans. R. Soc. Lond. B* **367**, 1338–1349 (2012).
5. Yeung, N. & Summerfield, C. *Phil. Trans. R. Soc. Lond. B* **367**, 1310–1321 (2012).
6. Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E. & Lau, H. *Cogn. Neurosci.* **1**, 165–175 (2010).
7. Pannu, J. K. & Kaszniak, A. W. *Neuropsychol. Rev.* **15**, 105–130 (2005).
8. Fleming, S. M. et al. *Psychol. Sci.* **26**, 89–98 (2015).
9. Shadlen, M. N. & Kiani, R. *Neuron* **80**, 791–806 (2013).
10. Fleming, S. M., Huijgen, J. & Dolan, R. J. *J. Neurosci.* **32**, 6117–6125 (2012).
11. Pleskac, T. J. & Busemeyer, J. R. *Psychol. Rev.* **117**, 864–901 (2010).
12. van den Berg, R. et al. *eLife* **5**, e12192 (2016).
13. Cavanagh, J. F. & Frank, M. J. *Trends Cogn. Sci.* **18**, 414–421 (2014).
14. Cohen, M. X. & Donner, T. H. *J. Neurophysiol.* **110**, 2752–2763 (2013).
15. Murphy, P. R., Robertson, I. H., Harty, S. & O'Connell, R. G. *eLife* https://doi.org/10.7554/eLife.11946 (2015).