Commentary on ‘How Something Can Be Said About Telling More Than We Can Know: On Choice Blindness and Introspective Report’.

Authors: James Moore* & Patrick Haggard

Address: Institute of Cognitive Neuroscience, 17 Queen Square, London, WC1N 3AR

* Author for correspondence. Email address: j.w.moore@ucl.ac.uk; Tel: 020 7679 1128

Acknowledgements: This work was supported by an ESRC/MRC studentship grant, and by ESRC grant RES00231571 to PH.

Everyday we offer ourselves explanations for the things we do and the choices we make, but how accurate are these introspections? This was a question famously tackled by Nisbett and Wilson in their seminal article: Telling more than we can know: Verbal reports on mental processes (1977). Their radical and counter-intuitive answer was that our introspections are confabulatory.

Despite the splash created by Nisbett and Wilson's article, and their proposed paradigm for testing their hypothesis, no coherent research programme emerged. This is a situation that Johansson and colleagues have sought to address with their 'Choice Blindness Paradigm' (CBP; see Johansson, Hall, Sikstrom, Tarning, & Lind, current issue).

In line with Nisbett and Wilson's hypothesis, the CBP suggests that our introspections are confabulatory. Johansson, Hall, Sikstrom, and Olsson (2005) presented participants with photographs of two female faces, one of which they had to choose as being more attractive. The 'chosen' photograph was then re-presented to the participant, who had to offer a justification for choosing that photograph. Unbeknownst to the participant, the experimenters intermittently swapped the photograph that was chosen, and instead presented the un-chosen one. Interestingly, Johansson et al found that when they presented to the participant a photograph they had not in fact chosen, participants would nevertheless offer a justification for that 'choice'.

This study appears to be a neat demonstration of Nisbett and Wilson's hypothesis. Participants clearly offered confabulatory explanations for choices they had not in fact

made. The strength of this study lies in the fact that one can more clearly discern the real from the confabulatory in these introspective reports. Moreover, Johansson, Hall, Sikstrom, Tarning, and Lind (this issue) reveal that real and confabulatory reports differ very little in terms of content. This finding is particularly telling. It implies that our justifications for 'real' choices may be based on the same processes that generate justifications for confabulatory choices.

A key issue is how far we should accept the conclusions of Johansson et al's study. Is it the case that all our introspections are detached from reality in this way? The psychological literature on the feature of voluntary action called 'agency' provides a domain where enough psychological data exist to address this concern.

Agency, broadly construed, is the ability to interact with the environment through self-generated action. Agency involves specific neural processes, their physical consequences in the environment, and also a characteristic conscious experience of action control. We can therefore ask if the conscious experience of agency is based on a confabulatory process of the sort posited by Johansson et al, or on genuine, specifiable information internal to the processes of action control.

Daniel Wegner and colleagues appear to suggest that introspections on agency are confabulatory. He writes '…we are not intrinsically informed of our own authorship and instead must build it up virtually out of perceptions of the thought and the actions we witness in consciousness (p. 218)'. Support for this assertion comes from a number of sources. Wegner and Wheatley (1999) showed that participants who were primed with an action-relevant thought prior to performing that action felt a

heightened sense of agency, even when they themselves did not perform that particular action. Furthermore, an erroneous sense of agency can occur in various clinical conditions. For example, patients with 'utilisation behaviour' will make well-formed actions directed at objects in their environment without consciously intending the action. They recognise the action is theirs, though they do not experience any intention to make it (Marcel, 2005). Although the action was not consciously intended, such patients will nevertheless offer post-hoc rationalisations for their actions. For example, Boccardi , Della Sala, Motto, and Spinnler (2002) provide the following example of a patient they tested with utilisation behaviour:

> '… while tested, CU spotted an apple and a knife left on purpose on a corner of the testing desk. He peeled the apple and ate it. The examiner asked why he was eating the apple. He replied "Well...it was there", "Are you hungry?" "No, well a bit", "Have you not just finished eating?" "Yes", "Is this apple yours?" "No". "And whose apple is it?" "Yours, I think", "So why are you eating it?" "Because it is here"' (p. 293).

These experimental and clinical examples appear to provide convincing evidence in support of the hypothesis of confabulatory introspection.

However, these are exceptions to the norm. For example, in Wegner and Wheatley's study, two agents participated in the experiment, and a given environmental effect could be caused either by one or by the other. Therefore the sense of agency was highly fallible. In the case of utilisation behaviour, there is severe lesioning to the

frontal lobes. In such cases, it may be the case that our sense of agency is indeed confabulatory, but only when intrinsic sources of information are made ambiguous (through the introduction of other possible causes as in Wegner & Wheatley's study), or when they are impaired (as in the case of utilisation behaviour). Bayne and Levy (2006) point out that the lengths one has to go to in order to render the sense of agency fallible demonstrate the reliability of the underlying mechanisms.

What direct evidence is there that the normal sense of agency is valid and reliable? A study by Fried et al (1991) suggests that our sense of agency may be generated by preparatory neural processes that also generate our voluntary actions. During a preoperative procedure, Fried and colleagues electrically stimulated the supplementary motor area of neurosurgical patients. At low current levels the patients reported having urges to make particular movements, and at higher levels they actually made the movements that they previously reported an urge to perform. This result suggests that the initial 'urge' is a normal accompaniment of the neural processes that generate action. If the sense of agency were a confabulation, it would presumably be triggered by sensory feedback of the action itself. Each action would then require a retrospective explanation. However, Fried et al.'s result suggests that an experience related to agency is present *before* any physical action has occurred. The sense of agency seems to be based on internal information generated by the neural mechanism that is responsible for the action. Fried et al's study argues against a confabulatory account of agency.

A computational model of motor control developed by Wolpert and colleagues (see Wolpert & Ghahramani, 2000, for a review) supports the assertion that our sense of

agency may be introspectively valid. On this view, the contents of conscious awareness may include predictions made by feed-forward models within the motor control system (Blakemore, Wolpert, & Frith, 2002). This could also explain the Fried et al findings above; the patients' conscious intentions to move appeared to be based on the same processes involved in the generation of the movement

A recent study by Moore and Haggard (in prep.) provides further support for the idea that our sense of agency is introspectively valid. Previous studies have shown that voluntary actions and their effects are perceived closer together in time than is actually the case (Haggard, Clark, & Kalogeras, 2002). This has been termed 'intentional binding'. Moore and Haggard used this finding to see whether the binding effect was dependent on the actual occurrence of the effect, or on the prediction that the effect will occur. By manipulating the predictability of the effect (a tone), we showed that, where predictability was high, actions showed a binding effect even in the absence of the tone. Where predictability of the effect was low, there was no such shift. To the extent that the binding phenomenon is taken as an aspect of the sense of agency, this finding suggests a predictive component to agency. The sense of agency appears to be based, at least in part, on predictions of the sensory consequences of our actions. Predictions are clearly not confabulations.

The picture emerging is that introspections are prone to confabulation where the sense of agency is fallible. However, when the sources of fallibility are removed, the internal information we have about our own agency is more reliable and more valid. Does CBP fall into the former cluster of cases in which the states we introspect on (in this case motivations for action) are artificially made fallible?

We suggest CBP *is* an aberrant case of this kind. For example, in the CBP the choice that is made is decidedly unimportant; it is unlikely that people profoundly care whether or not a face is attractive or not. Johansson's subjects could make sense of the trick situation in one of two ways. First, they could accept that the action that they made did not have the desired effect (showing the face that they had intended to choose). They would thus accept failed agency. Alternatively, they could confabulate new reasons for their action, which would retrospectively redefine their action as successful. In the artificial situation of the CBP experiment, confabulation is an easier method of 'sense-making' than accepting failed agency. A convincing refutation of this criticism would be a demonstration of the CBP effect for decisions regarding moral issues, for example. These would be decisions that are presumably less fallible and more resistant to confabulation.

Another key issue regarding the fallibility of introspection in the CBP is the experimenter-participant dynamic. There might be a feeling on behalf of the participant that whilst they suspect a mismatch between their intention and its effects, they are unwilling to admit as much to the experimenter. Again, this could be tested by getting participants to justify choices that are of a more important nature, or alternatively by giving participants independent evidence that their intentions will sometimes miscarry.

However, we should differentiate between access to one's *reasons* for performing an action, and access to the sense of agency itself (including intentions, authorship, conscious will, and so on). CBP appears to fall into the former class of cases, where

the task is to introspect on the reasons for a choice, not on the process of choosing itself. We suggest that confabulation about the *reasons* for acting is more common, whilst confabulations about the sense of agency itself are limited to unusual situations of ambiguity or impairment. We generally know about our own actions *when* we perform them, though we may be confused or self-deceptive about *why* we perform them. For example, in a situation of guilt, we commonly think of retrospective justifications or excuses for our action, while not denying that we performed it.

Whilst we welcome the introduction of the CBP as a useful experimental method, we suggest that caution should be exercised in the extent of its application. Undoubtedly there are many instances of confabulatory introspection. But confabulatory introspection does not work for all aspects of our action all the time. A key issue for future research is to try and better characterise the target of confabulation, and to differentiate normal access from exceptions. In general, we know about our own voluntary actions, before we make them. However, *reasons* for action seem to be more cognitively malleable, and susceptible to retrospective influences.

The idea that the true reasons for action may be hidden has a long history in psychology (Freud, 1923); we wish to suggest one possible explanation why reasons may be more malleable than agency. Agency often involves a direct phenomenal experience, of intention-in-action. We do not have direct phenomenal experience of *reasons* for action in the same way. Rather, our reasons for action, both predictive and retrospective, are based on the same general sense-making processes that we use to understand external events: the tree fell down because it was struck by lightning; I marked the examination because my boss said I had to; I bought flowers because I

knew it would make her happy. Systematic research on the processes which give us a sense of agency, and on the processes which give us reasons for action, is beginning, after a long post-behaviourist neglect. CBP will play an important part in this research, and we hope it can shed further light on the interaction between the experience of action and the thinking about reasons for action.

References

Bayne, T. & Levy, N. 2006. The Feeling of Doing: Deconstructing the Phenomenology of Agency. In N. Sebanz and W. Prinz (eds.) *Disorders of Volition*. Cambridge, MA: MIT Press.

Blakemore, S-J, Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences, 6(6),* 237-242.

Boccardi, E., Della Sala, S., Motto, C., & Spinnler, H. (2002). Utilisation behaviour consequent to bilateral SMA softening. *Cortex, 38,* 289-308.

Freud, S. (1923) 'The ego and the id'. In J. Strachey (ed.) *Standard Edition of the Complete Works of Sigmund Freud: Vol. 3.* New York: W. W. Norton.

Fried, I. Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., & Spencer, D. D. (1991). Functional organisation of human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience, 11,* 3656-3666.

Johansson, P., Hall, L., Sikstrom, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science, 310*, 116-119.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5(4)*, 382-385.

Marcel, A.J. (2005). The sense of agency: awareness and ownership of actions and intentions. In J. Roessler and N. Eilan, (eds.) *Agency and Self-Awareness*. Oxford University Press.

Moore, J. & Haggard, P. (in prep.). Predictive and inferential processes subserve the conscious awareness of goal-directed action.

Nisbett, T. D. & Wilson, R. E. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84(3)*, 231-257.

Wegner, D. M. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press

Wegner, D. M. & Wheatley, T. (1999). Apparent mental causation: sources of the experience of will. *American Psychologist, 54(7)*, 480-492

Wolpert, D. M. & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3*, 1212-1217.