Neuropsychologia 55 (2014) 137–142

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Action, agency and responsibility

Chris D. Frith [a,b,c,*]

[a] Wellcome Centre for Neuroimaging at University College London, 12 Queen Square, London WC1N 3BG, UK
[b] All Souls College, Oxford, UK
[c] Interacting Minds Centre, Aarhus University, Denmark

ARTICLE INFO

Available online 11 September 2013

Keywords:
Agency
Action
Will
Responsibility

ABSTRACT

In a series of experiments Marc Jeannerod revealed that we have very little awareness of the details and causes of our actions. We are, however, vividly aware of being in control of our actions and this gives us a sense of responsibility. These feelings arise, first, from intentional binding which creates a perception of agency, linking an intentional action to its outcome and, second, from the counterfactual reasoning that we could have chosen some other action. These feelings of responsibility play a critical role in creating social cohesion since they allow people to be held to account for deliberate antisocial behaviour. Jeannerod's studies also showed that we are unaware of how little we know about our actions and so are happy to make up stories about the nature and causes of our behaviour. These stories often do not correspond with the underlying cognitive and neural processes, but they can be changed through instructions and through discussion with others. Our experience of responsibility for action emerges during our upbringing through exposure to our culture. This creates consensus about the causes of behaviour, but not necessarily accuracy.

1. Introduction

Marc Jeannerod was a master in the design of experiments. In particular he had the ability to devise paradigms that seem on the surface very simple, but have the power to address deep problems about the mind. I am thinking, in particular, of his studies reaching and grasping (e.g., Castiello, Paulignan, & Jeannerod, 1991) and motor imagery (e.g., Jeannerod & Decety, 1995). In this essay I will consider just one of these elegant experiments and draw out some of its implications for our understanding of agency. I will also explore the relevance of our understanding of agency for the problem of free will. The problem of will was one of the Marc's major concerns throughout his career. The French subtitle of one of his early books was 'The physiology of will' (Jeannerod, 1983).

The particular problem I will address arises from the conflict between the vivid experience of being in control of our actions and the doctrine of determinism which concludes that this experience must be an illusion. I will explore the cultural origins of our experience of responsibility and suggest that the mechanisms that underlie this experience are compatible with determinism. Furthermore, our experience of responsibility plays a vital role in the maintenance of social cohesion.

2. The experience of action

The study reported by Fourneret and Jeannerod (1998) was the culmination of a decade long series of studies on the visual control of action (e.g., Pelisson, Prablanc, Goodale, & Jeannerod, 1986 see also O'Shea, et al., in press). In this study volunteers had to use their arm to move a stylus across a horizontal graphics tablet from a position close to the body to a target about 20 cm away. They could not see their arm, but only a cursor that was controlled by their arm movement. The trick in this study was that the movement of the cursor was sometimes distorted. When the stylus was moved straight ahead the cursor on the screen might deviate 10° to the right. Thus, in order to make the cursor move straight ahead, the volunteer had to make an arm movement that deviated 10° to the left. The striking result was that subjects seemed to be unaware of these distortions. They corrected their movements to allow for the distortions, but seemed unaware that they were making these corrections. For example, if they were asked to repeat the movement they had just made, then they made a straight-ahead movement and not the deviating movement they had just made, even in the absence of distracting visual feedback. The authors conclude, 'These results suggest that normal subjects are not aware of signals generated by their own movements'.

Many other experiments have confirmed this observation and shown that movements can occur, not only without awareness, but also without intentional control (e.g., Pisella et al., 2000). Such results led the philosopher, Metzinger (2006) to say that the

* Correspondence address: Wellcome Centre for Neuroimaging at University College London, 12 Queen Square, London WC1N 3BG, UK.
E-mail address: c.frith@ucl.ac.uk

0028-3932/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neuropsychologia.2013.09.007

phenomenology of volition, 'is thin and it is evasive'. So what is it about our actions that we are aware of?

While we are not aware of the details of our actions or the sensations associated with them, we are vividly aware of being in control. The doctrine of determinism, however, states that every event is necessitated by past history and conditions according to the laws of nature. From this point of view all our actions are also predetermined and so our experience of being in control is an illusion. The implication is that our belief in free will is also an illusion, a supposition supported by many, from Hume (1758) to Wegner (2003). Why then is our belief in free will so strong?

Some believe that it is our experience of being an agent that leads to a belief in free will. Spinoza (1677) said, 'Experience teaches us no less clearly than reason, that men believe themselves to be free, simply because they are conscious of their actions, and unconscious of the causes whereby those actions are determined'. Recent experimental studies are beginning to reveal, firstly, the cognitive basis of the experience of being an agent and, secondly, to hint as to why the belief that we are control of our actions may be important, even if it is an illusion.

However, before I consider these developments, I suggest that it might be more helpful to think about our actions in terms of responsibility rather than free will. Our culture makes an important distinction between voluntary and involuntary behaviour and also between outcomes that are intended and those that are not intended. We are not usually held responsible for acts that were involuntary or outcomes that were unintended. Free will and responsibility are closely related and many legal systems require that we can only be held fully responsible for those actions we have freely chosen. This is sometimes referred to as moral responsibility. However, the relationship between the moral and the legal is never perfect. Actions can be legal, while still being widely considered to be immoral. Nevertheless, the attribution of responsibility has an important function in society and has a major impact on our lives.

So what is it about our experience that makes us feel responsible for our actions? This is not a new question. Hellenistic philosophers, such as Epicurus, picked out two critical components. The first is the feeling of being an agent; it was I that caused the outcome (Τήν ἐξ ἑαυτοῦ αἰ τίαν; the cause from me), rather than someone or something else (Bobzien, 2006). The second is the feeling that I could have chosen otherwise. This is also the basis for the experience of regret. 'For it is on the grounds that it was possible for us also not to have chosen and not to have done this that we feel regret' (On Fate: Alexander of Aphrodisias, see Bobzien, 1998).

2.1. The feeling of agency

In a series of elegant experiments, Patrick Haggard and colleagues have demonstrated the phenomenon of *intentional binding,* which is probably fundamental to the experience agency (Haggard, Clark, & Kalogeras, 2002). In a typical experiment, participants press a button with a tone being heard 250 ms later. Using the paradigm introduce by Libet, Gleason, Wright, and Pearl (1983) participants are asked after each trial to indicate either when they pressed the button, or when they heard the tone. The times reported, which we might call *mental times*, are systematically shifted from the associated physical times. The time of pressing the button is perceived to be later, and the time of hearing the tone is perceived to be earlier. In other words, the interval between pressing the button and hearing the tone is smaller in mental time than in physical time. The action is bound closer to its outcome.

The critical result, however, is that this binding of action to outcome only occurs when the action is voluntary and deliberate. When the same finger movement is involuntary, having been caused by magnetic stimulation to the motor cortex, then the experienced times of the movement and the tone are pulled apart.

Timing is an example of a strong cue for causation. We perceive causation when an effect occurs shortly after a likely cause. And this perception can be an illusion. If we think of an outcome and then it occurs shortly afterwards we are likely to believe that we caused it even when we did not (Wegner & Wheatley, 1999). On the basis of our expectation before the action and on what occurs after the action the brain creates a perception of agency; of me causing the action. Intentional binding is an objective marker of this experience.

If responsibility is important for our sense of agency we would expect this experience to be enhanced when the outcome of our action has moral consequences. This expectation was confirmed in an experiment by Moretto, Walsh, and Haggard (2011) who found enhanced intentional binding when an action was followed by a moral rather than a merely economic outcome.

The other strong experience associated with agency is regret. Regret has to be distinguished from disappointment. We feel disappointment when the outcome of our action is worse than we expected. We feel regret when we learn that we would have achieved a better outcome if only we had chosen the other action. Regret is strongly linked to agency because, to feel regret, we have to believe that we could have chosen the other action. We feel much less regret if the action was chosen by the majority decision of a group, even though our decision was the same as the majority (Nicolle, Bach, Frith, & Dolan, 2011). We also feel less regret if we chose a habitual action rather than a novel one (Guttentag & Ferrell, 2004). In both the cases the degree of regret goes with the degree of agency.

Of particular interest is the observation that our decisions are affected by the anticipation of regret. We make our choice, not only on the basis of the anticipated outcomes associated with potential actions, but also on the regret we would expect to feel if it turns out that the option not chosen would have been better. This effect can be demonstrated by comparing situations in which the outcomes associated with the unchosen options will be known with situations where they will not be known. For example, when bidding in an auction we might be told that we had lost and also the amount of the winning bid or we might simply be told that we had lost. The former situation has much more potential for regret, if, for example, it turns out that the winning bid was only a little higher than our own. To avoid this anticipated regret people make higher bids when they expect to be told the winning bid (Filiz-Ozbay & Ozbay, 2007).

2.2. The experience of agency has social consequences

The success of social groups depends upon reciprocity and trust. We help others in the expectation that others will help us in the future. The maintenance of cooperation has been studied in the laboratory using trust games. Participants in these games can invest money in the group. The nature of the investment is such that, while each individual loses slightly from such investment, the group as a whole gains. So, if everyone invests, then everyone gains. There is, however, a strong temptation to get a free ride. That is, gaining more by benefitting from the investments of others while not investing oneself. Over time, the appearance of free riders lowers cooperation and everyone in the group suffers (Fehr and Gachter, 2002, #1026). Cooperation can be maintained and even enhanced; however, by allowing sanctions. Participants can pay small amounts to punish, through fines, other members of the group. This punishment is typically applied to free riders. As a result free riding decreases, cooperation is maintained and

everyone in the group benefits (Fehr and Gachter, 2002). Furthermore, in the long run, people will choose to be in a group where such sanctions are applied rather than in a group without sanctions (Gürerk, Irlenbusch, & Rockenbach, 2006).

The relevance of all this to the study of agency is the observation that such punishments are only applied to participants who are considered responsible for their actions. Singer et al. (2006) asked participants to award reward and punishments points to players in a trust game. As would be expected people who invested a majority of their stake were rewarded, while those who only invested a very small proportion were punished. However, the participants were told that some of the players were not freely deciding how much to invest, but were simply following written instructions. These players were neither punished, nor rewarded, even though their behaviour was identical to that of the players who made their decisions freely.

This result suggests that the concepts of agency and responsibility are important for determining whether people should be punished or rewarded for their behaviour, which, in turn, has a critical role in maintaining group cooperation and social cohesion. Of course, this idea is not new. The distinction between deliberate, intentional acts and outcomes that are unintended and accidental is at the basis of most legal systems. Furthermore, these systems assume that responsibility is something that emerges during childhood and is uniquely human. Non-human animals are no longer tried in court (Humphrey, 2002) and, below a certain age, children are not considered to be responsible for their actions. However, this minimum age of criminal responsibility varies widely from one legal system to another, even within Europe. For example, in Switzerland it is 7 years, while in Luxembourg it is 18 years (see Hazel, 2008). Some believe that even 18 is too young for full responsibility. The argument here is that the frontal lobes are still not fully mature in adolescence (see e.g., Mackintosh, 2011).

2.3. Responsibility and the brain

What are the neural mechanisms that underlie our experience of action? Why do we need mature frontal lobes to be considered fully responsible for our actions? I have suggested that there are two aspects of the experience of agency. The first is the feeling that I am the cause of some outcome. The second is that I could have done something else. Both of these experiences are critical to the concept of responsibility.

The feeling that I am the cause of an outcome allows us to distinguish between deliberately intended and accidental outcomes. This experience is a form of perception created by binding together expectations, based on intentions and prior beliefs, with sensations associated with the outcome (Moore & Haggard, 2008). But, in addition to the binding together of intention and outcome, we need to make a metacognitive reflection upon the extent to which we are in control of our actions (i.e. a judgement of agency). When people are asked to make judgements about their degree of agency, activity is seen in frontopolar cortex (BA 10, Miele, Wager, Mitchell, & Metcalfe, 2011). Other aspects of metacognition, such as reflecting on the accuracy of our perceptual judgements (confidence), have also been associated with processes associated with frontopolar cortex (Fleming, Weil, Nagy, Dolan, & Rees, 2010).

The feeling that I could have done something different, associated with regret, has also been associated with processes occurring in prefrontal cortex. Patients with lesions to orbital frontal cortex do not seem to experience regret and their behaviour is not influenced by anticipated regret (Camille et al., 2004). In an imaging study anticipation of regret was associated with activity in medial orbitofrontal cortex (Coricelli et al., 2005).

The feeling that I could have done something different is an example of counterfactual thinking. The most basic example of the influence of a counterfactual process on behaviour occurs when we learn from the actions we could have performed and not just from the ones we did perform. Consider the game stone–paper–scissors. I play *rock* and I am beaten by *paper*. If I am a simple reinforcement learning device, then I will lower my estimate of the value of playing *rock*. But I learn nothing new about the value of playing *paper* or *scissors*. However, if I use counterfactual thinking, then I can note that I would have won if I had played *scissors* and can increase my estimate of the value of playing this move. Updating of values on the basis of actions that could have been made has been observed in the behaviour of monkeys (Abe & Lee, 2011), as well as humans (Boorman, Behrens, Woolrich, & Rushworth, 2009), and is particularly associated with processes instantiated in frontopolar cortex (Platt & Hayden, 2011).

2.4. Responsibility, free will and determinism

Many neuroscientists (e.g., Greene & Cohen, 2004; Roth, 2010) believe that free will is incompatible with determinism and, therefore conclude that free will is an illusion. I have argued here that, even if it is an illusion, this feeling is critical for the acceptance that we are responsible for our actions and can be held to account for them. Thus this feeling of responsibility has a very important role in creating social cohesion. I suggest that the cognitive basis for the feeling of responsibility is, first, a mechanism that binds intentions to outcomes. This makes a distinction between expected and unexpected outcomes. Second, is the process that enables us to think about what would have happened if we had performed a different action, an example of counterfactual thinking? Both of these processes could occur in a fully determined manner, while still leading to the experiences relevant to the feeling of responsibility.

3. The cultural creation of responsibility

3.1. Stories about experience

I highlighted the experiment by Fourneret and Jeannerod (1998) because it suggests how little awareness we have of our own actions. There is another important lesson to be learned from this experiment. This is that we are not aware of how little insight we have into our own actions. We think we know all about our actions and are eager to explain and justify them. This practice applies, not only to the rather limited kinds of action we study in the lab, but also to the more important decisions that we make in real life.

In the Fourneret and Jeannerod experiment, participants were asked to repeat the action they had just made, without being given any visual feedback. Even though they had just moved their arm 10° to the left, they believed they had moved their arm straight ahead and it was the movement that they believed they had made that they reproduced. There are many more realistic examples of such failures of insight. Indeed, it has become a basic tenet of cognitive psychology that we have little or no knowledge of the cognitive processes that underlie our behaviour (Kihlstrom, 1987). Nevertheless, we are happy to give reports on these processes, which are based on intuitive theories about the causes of behaviour, sometimes called folk psychology (Nisbett & Wilson, 1977). Nisbett and Wilson report many cases of where these explanations of behaviour are erroneous. A widely quoted example concerns a position bias in choice behaviour (but see Kühberger, Kogler, Hug, & Mösl, 2006; Wilson & Nisbett, 1978). There was a pronounced bias to choose the rightmost item in an array of identical nylon stockings, but subjects never gave this as a reason for their choice

and, when the possibility was raised, denied that there was an effect of position.

A particularly striking example of our willingness to make up erroneous *post facto* justifications for behaviour is found in the phenomenon of change blindness (Hall, Johansson, Tarning, Sikstrom, & Deutgen, 2010; Johansson, Hall, Sikstrom, & Olsson, 2005). In one experiment participants were shown a series of pairs of faces and asked which one they preferred. They were then immediately shown again the face they had just chosen and asked to explain why they preferred it. On a small number of trials, by the use of sleight of hand, they were actually shown the face they had just rejected. Nevertheless, most of the time, participants would proceed to justify and explain this choice that they had not actually made.

That we are not aware of cognitive processes underlying choice behaviour is, perhaps, not so surprising. We would not expect to be aware of the neural activity that underlies these processes, so what is it that we could become aware of through introspection? Presumably the cognitive processes will generate some conscious experiences. For example, we seem to be aware of the time taken by a process to complete. This leads to the experiences such as *perceptual fluency* (the time taken to recognise something, Bornstein & D'Agostino, 1994) and *action selection fluency* (the time taken to choose an action, Chambon & Haggard, 2012). However, these experiences do not give us direct access to the underlying cognitive processes that generate them. We can use such experiences to make inferences about the processes, but these inferences can sometimes be wrong.

Consider the experience of perceptual fluency. This experience seems to have a role in the *mere exposure effect*. In this paradigm people are presented with a series of pictures, each presented very briefly. Afterwards they are unable to distinguish the pictures they have seen before from novel pictures. However, if asked which pictures they prefer, they tend to choose the ones they have seen before (Kunst-Wilson & Zajonc, 1980). This result shows that people have access to cues indicating which pictures have been seen before, but misinterpret these cues as indicating liking for a stimulus. One possible cue is perceptual fluency (Bornstein & D'Agostino, 1994). We can perceive a picture more quickly when we have seen it before and we are aware of differences in this perceptual fluency (Forster, Leder, & Ansorge, 2012). For reasons that need to be explored further we associate perceptual fluency with liking (Reber, Schwarz, & Winkielman, 2004).

A similar story can be told about the experience of emotion. In the well known experiment of Schachter and Singer (1962) subjects experienced an increase in physiological arousal caused by an injection of adrenaline. However, how they interpreted this experience in terms of emotion (euphoria or anger) depended on instructions and context. On the basis of these and other studies I agree with Schwitzgebel (2008) that our insight into our own experiences may be highly unreliable and provides very limited and inaccurate access to the underlying cognitive processes.

3.2. How discussion and instruction can change our experiences and behaviour

Although we have so little insight into the causes of our actions and the nature of our experiences, we are very happy to make up stories about them. In turn these stories will influence how others understand and experience the causes of action. Ultimately these stories are derived from beliefs about actions and experiences that we learn from others and which are part of our culture. And, of course, these beliefs can change. Even a brief discussion can change the way we describe our experiences and also change our behaviour. I suggest that it is precisely because our introspection is so unreliable that we can adopt new stories about the causes of our experiences and behaviour.

A striking example of how stories about the causes of action can affect our behaviour, comes from a study of will power. If people have to sit next to a plate of nice food, but have been told not to eat any, then much mental effort is exerted in the attempt to resist temptation. After such exertion, people show less persistence. For example, they will give up more quickly when given impossible puzzles to solve (Baumeister, Bratslavsky, Muraven, & Tice, 1998). It seems that will power is a cognitive resource that can be depleted by exertion (ego depletion). More recently this idea has been questioned. Ego depletion experiments were repeated, but this time, participants were given different stories about the effects of exerting will power. One group was told, 'Working on a strenuous mental task can make you feel tired such that you need a break before accomplishing a new task', while the other group was told, 'Sometimes, working on a strenuous mental task can make you feel energised for further challenging activities'. The behaviour of the people in these groups corresponded to their instructions, with the 'energised' group making fewer errors on a STROOP task after their strenuous mental activity, while the 'depleted' group made more errors (Job, Dweck, & Walton, 2010).

Much further work is needed to explore the mechanisms underlying these effects. One possibility is that, through discussion, people learn to reinterpret the fragile cues associated with the experience of decision-making and action. I already discussed the *mere exposure effect* in which people mistake the experience of perceptual fluency as a cue of liking rather than of prior exposure. My prediction would be that it should be possible to teach people to use the feeling of perceptual fluency as a cue that they have seen the picture before, thus making their story about their experience closer to reality.

Discussion of *post facto* justifications of choice can also affect our perception at a much more basic level. Bahador Bahrami and colleagues have reported a series of experiments in social psychophysics. In these studies, two participants work together to detect signals in standard psychophysical paradigms. After making individual decisions, the pair discussed disagreements about the signal and came up with a consensus through discussion. As long as their perpetual abilities are roughly similar, then the pair achieves better performance than the better partner working alone (Bahrami et al., 2010). This advantage occurs over and above any individual learning. The group advantage can be modelled on the assumption that people share with each other their confidence in their decision about the signal and choose the answer of the more confident member on a trial by trial basis. This was confirmed when a detailed analysis was performed on what was actually said during the discussion (Fusaroli et al., 2012). Pairs developed, from scratch, linguistic tools for communicating and calibrating levels of confidence. Of particular relevance here is the observation that the individual performance of someone involved in such an interaction improved to a significantly greater extent than people who had just as much practice with the signal detection task, but worked on their own (Olsen et al., in preparation). Discussion of perceptual decision making with others can improve individual perception.

3.3. The emergence of responsibility and the experience of volition

I suggest that our experiences of volition and responsibility are also influenced by instruction and discussion. For example, the phenomenon of intentional binding can be influenced by instructions about the effects of action. Dogge, Schaap, Custers, Wegner, and Aarts (2012) increased the binding between an *involuntary* key press and a tone by encouraging participants 'to perceive themselves as the cause for producing the tone'.

Instructions can also decrease the influence of the subtle cues available in our experience of action. In difficult choice reaction time tasks people typically slow down after they have made an

error (post-error slowing). This is an example of metacognitive or intentional control. Performance is monitored and, when there is evidence of an error, behaviour is adjusted to prevent its re-occurrence. Rigoni, Wilquin, Brass, and Burle (2013) gave people text to read claiming that 'scientists now recognise that free will is an illusion' or control texts in which free will was not mentioned. The people in the group who had been primed to have a weaker belief in free will showed reduced post-error slowing in a reaction time task. In cognitive terms free will is equated with intentional control, so that a loss in the belief in free will leads to a weakening of the motivation to apply intentional control.

There are now several experiments demonstrating how beliefs about free will can alter volitional behaviour (e.g., Baumeister, Masicampo, & Dewall, 2009; Vohs & Schooler, 2008). I suggest, however, that the response to such instructions will depend on what people believe to be the effects of a loss of intentional control. For example, I suspect that most people believe that intentional control is needed to overcome our selfish impulses. However, it may well be that, in many circumstances, our basic impulse is to be prosocial (see for example Valdesolo & DeSteno, 2008). The effects of instructions are likely to depend upon what people believe about the relevant cognitive processes, rather than the actual nature of those cognitive processes.

Given the increasing evidence of how readily volitional behaviour and experience can be altered by instructions and beliefs, it would be surprising if there were no effects of upbringing and culture. Epicurus believed that we acquire the idea that we are causal agents through the observation that human beings, including ourselves, are praised and blamed for their actions (Bobzien, 2006). It is certainly the case that experiences of volition and responsibility arise rather slowly during the course of childhood.

I have suggested that the idea and experience of responsibility depend upon two factors. The first is the experience of being an agent and causing things to happen. The second is the counterfactual thought that I could have acted differently. It is this second factor that leads to the experience of regret. It is only by the age of 5 that children can report a mature experience of agency, distinguishing between a voluntary movement of the leg and a knee jerk reflex (Shultz, Wells, & Sarda, 1980). A true experience of regret seems to arise much later, at around the age of 9 (Rafetseder & Perner, 2012). Prior to this age children base their judgments solely on what they got without taking into account what they could have got. Finally there is some preliminary evidence that the excessive risk taking behaviour seen in adolescence may be related to a lack of, or failure to take account of, anticipated regret (Gerrard, Gibbons, Benthin, & Hessling, 1996). Thus it seems that the ability to take regret into consideration, when making choices, continues to develop during adolescence (Habib et al., 2012).

Are there cultural differences in the experience of agency? Among the Mopan Mayas of Central America punishment is applied equally if the outcome is accidental, rather than deliberate. Saying a falsehood is not excused even if, at the time, the speaker believes it to be true. Children who indulge in pretend play are reproved. It seems that, for this cultural group, the intention behind the action has no importance. It is only the outcome that counts in applying praise or blame (Danziger, 2006). It would be interesting and informative to study intentional binding and post-error slowing in such a group.

4. Conclusions

The studies of action that Marc Jeannerod pioneered concern behaviour that is far more restricted than actions occurring outside the laboratory. Nevertheless, I believe that what these studies reveal about our experience of action have important implications for behaviour in 'real life'. They reveal that our introspection is a very unreliable method for uncovering the cognitive basis of our actions. However, we like to reflect upon our behaviour and experiences and to discuss these reflections with others. Through these discussions we develop explicit accounts of the world and ourselves. It is the history of these discussions that largely determine individual and cultural differences.

The effects of culture on experience need not necessarily lead to a more accurate understanding of our actions. For example, our experience of being a causal agent may not accurately reflect the cognitive and neural processes underlying our actions. The important aspect of this story about agency is that we have a social consensus about the causes of our actions (Proust, 2012). Our experience of being responsible for our actions is extremely important for generating social cohesion and benefitting the members of the group. We all benefit from reciprocal altruism and, to the extent to which we can exert intentional control, from deliberately following explicit advice from other members of our group.

However, once social cohesion is established, it will enhance the effects of the various unconscious mechanisms I have discussed. Our desire to be part of the group increases our tendency to imitate others (e.g., Over & Carpenter, 2009) and to behave in a prosocial, rather than a selfish manner (e.g., Bateson, Nettle, & Roberts, 2006). Furthermore, this unconscious imitation of the behaviour of others takes advantage of the all that dangerous trial and error learning undertaken by previous generations (see Frith & Frith, 2012).

These powerful effects indicate the importance of our experience of being responsible agents with intentional control. But is this experience merely a necessary illusion or do we really have some degree of intentional control over our actions? Reflecting on the regret we might feel, if it turns out that we have chosen the wrong option, can affect our behaviour (e.g., Filiz-Ozbay & Ozbay, 2007). Furthermore, the reflection on our actions engendered by discussions with others, and by instructions, can also change our behaviour (e.g., Rigoni et al., 2013). I believe these are demonstrations that we do sometimes have intentional control over our behaviour. We are not merely automatons in a deterministic world. Through counter-factual thinking we can create what might have been and what might be. We are not limited to reacting to the world. We can also change it (Friston, Daunizeau, Kilner, & Kiebel, 2010).

These reflections on responsibility reveal Marc Jeannerod's legacy to me in relation to my understanding of the nature of volition. As revealed by the study of Fourneret and Jeannerod (1998) and many of Marc's other studies, our awareness of our actions is very limited and very fragile. This fragility can lead to incorrect interpretations. But these interpretations can be altered by discussion.

Marc's discoveries about the nature of action have opened a window onto the mechanisms underlying human interactions and the emergence of culture.

Acknowledgements

This work was supported by the Wellcome Centre for Neuroimaging at UCL and All Souls College, Oxford. I am grateful to Uta Frith & Rosalind Ridley for their comments on earlier drafts of this paper and to Panagiotis Mitkidis for checking my Greek. The paper has, I hope, been much improved by revisions in response to the comments of two anonymous referees.

References

Abe, H., & Lee, D. (2011). Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron*, *70*, 731–741.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*, 1081–1085.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*, 412–414.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*, 1252–1265.

Baumeister, R. F., Masicampo, E. J., & Dewall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, *35*, 260–268.

Bobzien, S. (1998). *Determinism and freedom in stoic philosophy*. Oxford: Clarendon Press.

Bobzien, S. (2006). Moral responsibility and moral development in epicurus' philosophy. In: B. Reis (Ed.), *The virtuous life in Greek ethics* (pp. 206–299). New York: Cambridge University Press.

Boorman, E. D., Behrens, T. E., Woolrich, M. W., & Rushworth, M. F. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, *62*, 733–743.

Bornstein, R. F., & D'Agostino, P. R. (1994). The Attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition*, *12*, 103–128.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J. R., & Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, *304*, 1167–1170.

Castiello, U., Paulignan, Y., & Jeannerod, M. (1991). Temporal dissociation of motor responses and subjective awareness. A study in normal subjects. *Brain*, *114*(Pt 6), 2639–2655.

Chambon, V., & Haggard, P. (2012). Sense of control depends on fluency of action selection, not motor performance. *Cognition*, *125*, 441–451.

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience*, *8*, 1255–1262.

Danziger, E. (2006). The thought that counts: Interactional consequences of variation in cultural theories of meaning. In: S. Levinson, & N. Enfield (Eds.), *The roots of human sociality: Culture, cognition and human interaction* (pp. 259–278). New York: Berg Press (Wenner-Gren Foundation for Anthropological Research).

Dogge, M., Schaap, M., Custers, R., Wegner, D. M., & Aarts, H. (2012). When moving without volition: Implied self-causation enhances binding strength between involuntary actions and effects. *Consciousness and Cognition*, *21*, 501–506.

Fehr, E., & Gachter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Filiz-Ozbay, E., & Ozbay, E. Y. (2007). Auctions with anticipated regret: Theory and experiment. *American Economic Review*, *97*, 1407–1418.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*, 1541–1543.

Forster, M., Leder, H., & Ansorge, U. (2012). It felt fluent, and I liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion*, *13*, 280–289.

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, *36*, 1133–1140.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biology Cybernatics*, *102*, 227–260.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, *63*, 287–313.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylen, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*, 931–939.

Gerrard, M., Gibbons, F. X., Benthin, A. C., & Hessling, R. M. (1996). A longitudinal study of the reciprocal nature of risk behaviors and cognitions in adolescents: What you do shapes what you think, and vice versa. *Health Psychology*, *15*, 344–354.

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, *359*, 1775–1785.

Guttentag, R., & Ferrell, J. (2004). Reality compared with its alternatives: Age differences in judgments of regret and relief. *Developmental Psychology*, *40*, 764–775.

Gürerk, O., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, *312*, 108–111.

Habib, M., Cassotti, M., Borst, G., Simon, G., Pineau, A., Houde, O., & Moutier, S. (2012). Counterfactually mediated emotions: A developmental study of regret and relief in a probabilistic gambling task. *Journal of Experimental Child Psychology*, *112*, 265–274.

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

Hall, L., Johansson, P., Tarning, B., Sikstrom, S., & Deutgen, T. (2010). Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea. *Cognition*, *117*, 54–61.

Hazel, N. (2008). *Cross-national comparison of youth justice* (p. 74). London: Youth Justice Board74.

Hume, D. (1758). *An enquiry concerning human understanding* (p. 2005)Oxford: Oxford University Press2005.

Humphrey, N. (2002). *Bugs and beast before the law. The mind made flesh* (pp. 235–254)Oxford: Oxford University Press235–254.

Jeannerod, M. (1983). *Le cerveau–machine. Physiologie de la volonté*. Paris: Librarie Arthème Fayard.

Jeannerod, M., & Decety, J. (1995). Mental motor imagery: A window into the representational stages of action. *Current Opinion in Neurobiology*, *5*, 727–732.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion—is it all in your head?: Implicit theories about willpower affect self-regulation. *Psychological Science*

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.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science*, *237*, 1445–1452.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*, 557–558.

Kühberger, A., Kogler, C., Hug, A., & Mösl, E. (2006). The role of the position effect in theory and simulation. *Mind and Language*, *21*, 610–625.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, *106*(Pt 3), 623–642.

Mackintosh, N. (2011). *Brain waves module 4: Neuroscience and the law* (p. 35) London: Royal Society35.

Metzinger, T. (2006). Conscious volition and mental representation: Toward a more fine-grained analysis. In: N. Sebanz, & W. Prinz (Eds.), *Disorders of volition* (pp. 19–48). Cambridge, Mass: Bradford Books, MIT Press.

Miele, D. B., Wager, T. D., Mitchell, J. P., & Metcalfe, J. (2011). Dissociating neural correlates of action monitoring and metacognition of agency. *Journal of Cognitive Neuroscience*, *23*, 3620–3636.

Moore, J., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Conscious and Cognition*, *17*, 136–144.

Moretto, G., Walsh, E., & Haggard, P. (2011). Experience of agency and sense of responsibility. *Conscious and Cognition*, *20*, 1847–1854.

Nicolle, A., Bach, D. R., Frith, C., & Dolan, R. J. (2011). Amygdala involvement in self-blame regret. *Society for Neuroscience*, *6*, 178–189.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know—verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Olsen, K., Bahrami, B., Christensen, P., Rees, G., Roepstorff, A., & Frith, C. D. 2013Human interaction accelerates visual perceptual learning in individuals, (in preparation).

O'Shea, J., Gaveau, V., Kandel, M., Koga, K., Susami, K., Prablanc, C., & Rossetti, Y. (2013). Kinematic markers dissociate error correction from sensorimotor realignment during prism adaptation. *Neuropsychologia* (in press).

Over, H., & Carpenter, M. (2009). Priming third-party ostracism increases affiliative imitation in children. *Developmental Science*, *12*, F1–F8.

Pelisson, D., Prablanc, C., Goodale, M. A., & Jeannerod, M. (1986). Visual control of reaching movements without vision of the limb. II. Evidence of fast unconscious processes correcting the trajectory of the hand to the final position of a double-step stimulus. *Experimental Brain Research*, *62*, 303–311.

Pisella, L., Grea, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., Boisson, D., & Rossetti, Y. (2000). An 'automatic pilot' for the hand in human posterior parietal cortex: Toward reinterpreting optic ataxia. *Nature Neuroscience*, *3*, 729–736.

Platt, M. L., & Hayden, B. (2011). Learning: Not just the facts, ma'am, but the counterfactuals as well. *PLoS Biology*, *9*, e1001092.

Rafetseder, E., & Perner, J. (2012). When the alternative would have been better: Counterfactual reasoning and the emergence of regret. *Cognition and Emotion*, *26*, 800–819.

Reber, R., Schwarz, N., & Winkielman, P. (2004). *Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?*. *Personality and Social Psychology Review*, *8*364–382.

Rigoni, D., Wilquin, H., Brass, M., & Burle, B. (2013). When errors do not matter: Weakening belief in intentional control impairs cognitive reaction to errors. *Cognition*, *127*, 264–269.

Roth, G. (2010). Free will—insights from neurobiology. In: U. J. Frey, C. Stormer, & K. P. Willfuhr (Eds.), *Homo novus—a human without illusions* (pp. 231–245). New York: Springer.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*, 379–399.

Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, *117*, 245–273.

Shultz, T. R., Wells, D., & Sarda, M. (1980). Development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *British Journal of Social and Clinical Psychology*, *19*, 301–310.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*, 466–469.

Spinoza, B. (1677). Part III: Of the affects. In *Ethics* (translation 1994) (p. 73). Penguin: London.

Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, *44*, 1334–1338.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, *19*, 49–54.

Wegner, D. M. (2003). *The illusion of conscious will*. Cambridge, Mass.: The MIT Press.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation—Sources of the experience of will. *American Psychologist*, *54*, 480–492.

Wilson, T. d. C., & Nisbett, R. E. (1978). The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. *Social Psychology*, *41*, 118–131.