

1

Neuroscience, choice and responsibility

PATRICIA S. CHURCHLAND

1.1 Introduction

Much of human social life depends on the notion that agents have control over their actions and are responsible for their choices. In daily life it is commonly assumed that it is fair to punish and reward behavior so long as the person is in control and makes choices knowingly and intentionally. Without the assumptions of agent control and responsibility, human social commerce is hardly conceivable. As members of a social species, we recognize co-operation, loyalty, honesty, and helping as prominent features of the social environment. We react with hostility when group members disappoint certain socially significant expectations. Inflicting disutilities (e.g., shunning, pinching) on the socially erring and rewarding civic virtue help restore the standards.

In other social species too, social unreliability, such as a failure to reciprocate grooming or food-sharing, provokes a reaction likely to cost the erring agent, sooner or later. In social mammals at least, mechanisms for learning and keeping the social order seem to be part of what evolution has bequeathed to our brain circuitry. Given that the stability of the social-expectation baseline is sufficiently important for survival, individuals are prepared to incur some cost in enforcing those expectations. Just as anubis baboons learn that tasty scorpions are to be found under rocks but cannot just be picked up, so they learn that failure to reciprocate grooming when it is duly expected may incur a slap.

In social species, parents invest heavily not only in the production, feeding, and protection of the young, but also in their socialization. The young of the social species must learn how to navigate the physical world, but their survival and flourishing also depends on their acquiring appropriate habits to navigate the social world. In both cases, the reward system plays a crucial role, enabling the brain to improve its predictions about what will satisfy, and what will bring pain.

Topics in Integrative Neuroscience: From Cells to Cognition, ed. James R. Pomerantz. Published by Cambridge University Press. © Cambridge University Press 2008.

If the reward and punishment system is to be effectively engaged in shaping social behavior, the actions for which the agent is rewarded or punished must be under the agent's control; that is, reward and punishment should make some difference to the agent's future predictions of an action's consequences and hence of his behavior. What does it mean, for us to have control over our behavior? Are we ever *really* responsible for our choices and decisions? Will neuroscientific understanding of the neuronal mechanisms for decision-making change how we think about these fundamental features of social commerce? These are the places where issues about free will bump up against practical reality of negotiating of what is fair, what is reasonable, and what is effective.

1.2 Are we in control if our choices and actions are *caused*?

One tradition bases the conditions for free will and control on a contrast between being *caused* to do something and *not* being so caused. For example, if someone falls on me and I hit you, then my hitting you was caused by the falling body; I did not choose to hit you. I am not, therefore, responsible for hitting you. Were you to punish me for hitting you, it would not help me avoid such events in the future. Examples emanating from this prototype have been extended to the broader idea that in order for *any* choice to be free, it must be absolutely *uncaused*; that is, it is suggested that a free choice is made when, without any prior cause and constraints, a decision comes into being and an action results. An example allegedly illustrating this idea is Eisenhower's decision to send troops into Little Rock to enforce school desegregation. Or my decision to go to the coffee shop for a cappuccino. This *contra-causal* construal of free choice is known as "libertarianism."¹ Is it plausible? That is, are the paradigm cases of free choices actually *uncaused* choices?

As Hume proposed in 1739,² the answer is *no*. Hume argued that our choices and decisions are in fact caused by other events in the mind – desires, beliefs, preferences, feelings, and so forth. Thus Eisenhower's decision was the outcome of his beliefs about the situation, and his desire to ensure that the federal school integration law was not flaunted. His decision did not suddenly and without preceding beliefs, thoughts, hopes, and worries spring uncaused into existence. I went to get a cappuccino because I usually have one about this time in the afternoon, I wanted to have one, and I knew I had enough money to pay for it, and so on. Save for these causal antecedents, albeit *cognitive* causal antecedents, I would not have gone for coffee. By contrast, suppose that without any antecedent causes, I suddenly enter a saloon, ask for a glass of vodka, and swig it back. I had no antecedent desire for vodka, no habit of going to a saloon anytime, let alone in the afternoon, and the behavior would be considered utterly at

4 Patricia S. Churchland

odds with my cognitive state and temperament. Is *this* the paradigm of free choice? Is *this* prototypically responsible behavior?

Reflecting on these sorts of possibilities, Hume made the deeper and more penetrating observation that an agent's choices are not considered freely made *unless* they are caused by his desires, intentions, and so forth. Randomness, pure chance, and utter unpredictability are not preconditions for attribution of responsible choice. Hume puts the issue with memorable compactness: "Where [actions] proceed not from some cause in the characters and disposition of the person, who perform'd them, they infix not themselves upon him, and can neither redound to his honor if good, nor infamy, if evil."³

Logic reveals, Hume argued, that responsible choice is actually *inconsistent* with libertarianism (uncaused choice). Someone may choose to climb onto his roof because he does not want the rain to come into his house, he wants to fix the loose shingles that allowed the rain in, and he believes that he needs to get up on the roof to do that. His desires, intentions, and beliefs are part of the causal antecedents resulting in his choice, though he may not be introspectively aware of them *as* causes. If, without any determining desires and beliefs, he simply went up onto the roof – as it were, *for no reason* – his sanity and hence his control is seriously in doubt.

More generally, a choice undetermined by anything the agent believes, intends, or desires is the kind of thing we consider *out* of the agent's control, and is not the sort of thing for which we hold someone responsible. Furthermore, desires or beliefs that are uncaused (assuming that is physically possible) rather than caused by other stable features of the person's character and temperament are likewise useless in producing the conditions for responsible choice. If a powerful desire with no antecedent connection to my other desires or my general character were to spring into my mind – say, the serious desire to become a seamstress – I would suspect that someone must be "messing with my mind." The brain presumably has no mechanism for introspectively recognizing a desire to fix the roof *as* a cause, just as it has no way of detecting through introspection that growth hormone has been released or that blood pressure is at 110/85. A cause, nevertheless, a desire most certainly is.

Neither Hume's argument that choices are internally caused nor his argument showing that libertarianism is absurd have ever been convincingly refuted. Notice, moreover, that his arguments hold regardless of whether the mind is a separate *Cartesian substance*, or a pattern of activity of the physical brain. And they hold regardless of whether the etiologically relevant states are conscious or unconscious.

In fact, the brain does indeed appear to be a causal machine. So far, there is no evidence at all that events at the neuronal or network or systems levels happen

without any cause. True enough, neuroscience is still in its early stages, and we cannot absolutely rule out the possibility that evidence will be forthcoming at some later stage. Given the data, however, the odds are against it. Importantly, even were uncaused neuronal events to be discovered, it is a *further* and substantial matter to show that precisely *those* neural events constitute free choice. They might, for all we can tell now, have to do with features of growth hormone release or variations in the sleep–wake cycle.

Nonetheless, the idea that randomness in the physical world is somehow the key to what makes free choice *free* remains appealing to those inclined to believe that free choice must be uncaused choice. With the advent of quantum mechanics and the respectability of the idea of quantum indeterminacy (i.e. the physical description of a state is necessarily incomplete), the suggestion that somehow or other quantum-level indeterminacy is the basis for a “solution” to the problem of free will remains attractive to some libertarians.⁴ Stripped to essentials, the hypothesis claims that although an agent may have the relevant desires, beliefs, etc., he still can make a choice that is truly independent of all antecedent causal conditions. On this view, the agent, not the agent’s brain or his desires or his emotions, freely chooses between cappuccino and latte, for example. It is at the moment of deciding that the indeterminacy or the non-causality or the break in the causal nexus – whatever one wants to call it – occurs. The subsequent choice is therefore absolutely free.

This is meant to be an empirical hypothesis and, as such, it needs to confront neurobiologically informed questions. For example, what exactly, in neural terms, is the *agent who chooses*? How does that fit with what we understand about self and self-representational capacities in the brain? Under exactly what conditions do the supposed noncaused events occur? Does noncausal choice exist only when I am dithering or agonizing between two equally good – or perhaps equally bad – alternatives? What about when, in full conversation, I use the word “very” rather than the word “extremely”? Does it exist with respect to the *generation* of desires? Why not? There are also questions from quantum physics such as these: What is the mechanism of amplification of the non-deterministic events? Were quantum effects of the envisioned kind to exist in the brain, how could they fail to be swamped by thermal indeterminacy?

These are just the first snowballs in an avalanche of empirically informed questions. Part of their effect is to unmask the flagrantly ad hoc character of the hypothesis; that is, it is based more on a desire to prop up a wobbling ideology than on factual matters. Rather than fully discussing its merits and flaws now, however, we shall defer a closer analysis of the hypothesis of a quantum-level origin for uncaused choice until further details of the neurobiology of decision-making are on the table. That will allow us to see what bearing the neurobiological data have on the question of causality and choice in the brain, and hence provide a richer

6 Patricia S. Churchland

context for evaluating the hypothesis of noncausal choice. Therefore, we return to this hypothesis and its critics in Section 1.6 to see how it fares.

Provisionally, therefore, let us adopt the competing hypothesis, namely that Hume is essentially right, and all choices and all behavior *are* caused, one way or another. The absolutely critical point, however, is that not all causes are equal before the tribunal of responsibility: some are such as to excuse us from culpability; others precisely render us culpable. The important question concerns the relevant differences among causes of behavior such that some kinds play a role in free choice and others play a role in forced choice; that is: Are there systematic *brain-based* differences between voluntary and involuntary actions that will support the notion of agent responsibility? This is the crucial question, because we do hold people responsible for what we take to be *their* actions. When those actions are intentionally harmful to others, punishment, varying from social disapproval to execution, may be visited upon the agent. When, if ever, is it fair to hold an agent responsible? When, if ever, is punishment justified?

Many possibilities have been explored to explain how the notions of control and responsibility can make sense in the context of causation. These fall under the general rubric, “Compatibilism,” meaning that our work-a-day notion of responsibility is, at bottom, *compatible* with the probable truth that the mind-brain is a causal machine. First we shall consider some obvious but unsuccessful attempts, and then we shall raise the possibility that increased understanding of the brain will aid in piecing together a plausible account.⁵

1.3 Caused choice and free choice: some traditional hypotheses

1.3.1 *Voluntary causes are internal causes*

Can we rely on this rule: you are responsible if the causes are internal, otherwise not? No, for several reasons. A patient with Huntington’s disease makes nonpurposeful, jerky movements as a result of internal causes. But we do not hold the Huntington’s patient responsible for his movements, since they are the outcome of a disease that causes destruction in the striatum. He has no control over his movements, and they are not consistent with his actual desires and intentions, which he cannot execute. A sleepwalker may unplug the phone or kick the dog. Here too the causes are internal, but the sleepwalker is not straightforwardly responsible. In a rather attenuated sense, the sleepwalker may *intend* his movements, though he is apparently unaware of his intentions.

1.3.2 *Voluntary causes are internal, they involve the agent’s intentions, and the agent must be aware of the intention*

This strategy also fails. A patient with obsessive-compulsive disorder (OCD) may have an overwhelming urge to wash her hands. She wants and

intends to wash her hands, and she is fully aware of her desire and her intention. She knows that the desire is her desire; she knows that it is she who is washing her hands. Nevertheless, in patients with OCD, obsessive behavior such as hand-washing or footstep-counting is considered to be out of the agent's control. They often indicate that they wish to be rid of hand-washing or footstep-counting behavior, but cannot stop. Pharmacological interventions, such as Prozac, may enable the subject to have what we would all regard as normal choice about whether or not to wash her hands.

1.3.3 *Voluntary causes feel different from the inside*

Another strategy is to base the distinction on *felt* differences in inner experience between those actions we choose to do and those over which we feel we have no control. Thus it allegedly feels different when we evince a cry as a startle response to a mouse leaping out of the compost heap, and when we cry out to get someone's attention and help. Is introspection a reliable guide to responsibility? Can introspection – attentive, careful, knowledgeable introspection – distinguish those internal causes for which we are responsible from those for which we are not? (See also Crick 1994.)

Probably not. There are undoubtedly many cases where introspection is no guide at all. Phobic patients, the OCD patients just mentioned, and those with Tourette's syndrome are obvious examples that muddy the waters. In a patient with claustrophobia, the desire not to go into a cave feels as much *his* as his desire not to go rafting without a life jacket. He can even give reasons for both – it could be unsafe, avoidable injuries could happen, etc. His desire not to go into a cave may be very very strong, but so may be his desire to eat when hungry or sleep with his wife. So mere *strength* of desire will not suffice to distinguish actions for which the agent has undiminished responsibility and those for which he is not fully responsible.

The various kinds of addictions present a further range of difficulties. A smoker feels that the desire for a cigarette is indeed *his*. His reaching for a cigarette may feel every bit as free as reaching to turn on the television or scratching his nose. He might wish it were not his, but so far as the *feeling* itself is concerned, it is as much his as his desire to quit smoking. The increase in intensity of sexual interest and desire at puberty is surely the result of hormonal changes on the brain, not something over which one has much control. Yet all of that interest, inclination, and alteration of behavior *feels* – from the inside at any rate – entirely free.

More problematic, perhaps, are the many examples from everyday life where one may suppose the decision was entirely one's own, only to discover that subtle manipulation of desires by others had in fact been the decisive factor.

8 Patricia S. Churchland

According to the fashion standards of the day, one finds certain clothes beautiful, others frumpy, and the choice of wardrobe seems, introspectively, as free as any choice. There is no escaping, however, the fact that what is in fashion has a huge effect on what we find beautiful, and this affects not only choice in clothes, but also such things as aesthetic judgment regarding plumpness or slenderness of the female body. Baseball hats worn backward have been in fashion for about ten years and are considered to look good, but from another perspective, most people look less attractive if wearing a backward baseball cap.

Social psychologists have produced dozens of examples that further muddy the waters, but a simple one will convey the point. On a table in a shopping mall, experimenters placed ten pairs of identical panty hose, and asked shoppers to select a pair, and then briefly explain their choice. Choosers referred to color, denier, sheerness, and so forth as their rationale. In fact, there was a huge position effect: shoppers tended to pick the panty hose in the rightmost position on the table. None of them considered this to be a factor, none of them referred to it as a basis for choice, yet it clearly was so. The ten pairs of panty hose were, after all, identical to one another. Other examples of priming, subliminal perception, and emotional manipulation also suggest that an appeal to introspection to solve our problem, about which behavior is in our control and which is not, is unlikely to go very far.⁶

1.3.4 *Could have done otherwise*

In a different attack on the problem, philosophers have explored the idea that if the choice were free, the agent *could have chosen otherwise*; that is, in some sense, the agent had the power to do something else.⁷ Certainly this idea does comport with conventional expectations about voluntary behavior, and insofar as it is appealing. Lyndon Johnson, historians say, could have done otherwise regarding Vietnam. He could have decided to stop the war in Vietnam in 1965 when he correctly judged it to be unwinnable. I could have decided not to get coffee, and perhaps to have water instead. Nobody *forced* me or coerced me; the desire for coffee was mine. So far so good. The weakness in the strategy shows up when we ask further, “what exactly does *that* mean?” If all behavior has antecedent causes, then “could have done otherwise” seems to boil down to “would have done otherwise if antecedent conditions had been different.” Accepting that equivalence means the criterion is too *weak* to distinguish between the shouted insults of a Touretteur, whose utterances include such random and undirected outbursts as “idiot, idiot, idiot,” and those of a member of parliament responding to an honorable member’s proposal, “idiot, idiot, idiot.” In both cases, had the antecedent conditions been different, obviously the results would have been different. Nevertheless, we hold the

parliamentarian responsible, but not the Touretter. So the proposed criterion seems not so much wrong, as unhelpful in revealing the nature of the difference between the causes of voluntary behavior and the causes of *nonvoluntary* behavior.

The further problem here is a lurking circularity. Testing for whether an agent could have done otherwise seems exactly the same as testing whether the behavior was voluntary. Hence specifying what counts as voluntary behavior by referring to the possibility that the agent might have done otherwise just goes around in a small circle. It does not seem to get us anywhere.

1.4 Prototypes and responsibility

In our legal as well as our daily practice, the pattern is to accept certain prototypical conditions as excusing a person from responsibility, but to assume him responsible unless a definite exculpatory condition obtains. In other words, responsibility is the default condition; excuse from and mitigation of responsibility has to be positively established. The set of conditions regarded as exculpatory can be modified as we learn more about behavior and its etiology. A different but related issue concerns what to do with someone who harms others but has diminished responsibility.

Aristotle (384–322 BC) in his great work, *The Nicomachean Ethics*, was the first to articulate the responsible-unless-exculpating-reasons principle. The wisdom of the principle is still reflected in much of human practice, including current legal practice. In his systematic and profoundly sensible way, Aristotle pointed out that for an agent to be held responsible, it is a necessary condition (but not a sufficient condition) that the cause be internal to the agent. In addition, he characterized as involuntary, actions produced by coercion and actions produced in certain kinds of ignorance. As Aristotle well knew, however, no simple rule demarcates cases here. Clearly, some ignorance is not considered excusable, when it may be fairly judged that the agent *should* have known. Additionally, in some cases of coercion, the agent is expected to resist the pressure, given the nature of the situation. A captured soldier is supposed to resist giving information to the enemy. As Aristotle illustrates in his own discussion of such complexities, we seem to proceed to deal with these cases by judging their similarity to uncontroversial and well-worn prototypes, which is perhaps why precedent law is so useful.⁸

Increasingly, it seems unlikely that there is a *sharp* distinction – brain-based or otherwise – between the voluntary and the involuntary – between being in control and being out of control – either in terms of behavioral conditions or in terms of the underlying neurobiology. This does not imply that there is *no*

10 Patricia S. Churchland

distinction, but only that whatever the distinction, it is not sharp; that is, it is not like the distinction between having a valid California driver's license and *not* having one. It is rather more like categories with a prototype structure, for example being a good sled dog, being a navigable river, being a fertile valley. These sorts of categories are useful even though we cannot specify necessary and sufficient conditions for membership in the category. We teach the category by citing prototypical instances, along with contrasting prototypical *non*instances.

Once we consider *being in control* in this light, we instantly recognize the degrees and nuances typical of freedom of choice. An agent's decision to change television channels may be more unconstrained than his decision to pay for his child's college tuition, which may be more unconstrained than his decision to marry his wife, which may be more unconstrained than his decision to turn off the alarm clock. Some desires or fears may be very powerful, others less so, and we may have more self-control in some circumstances than in others. Prolonged sleep deprivation makes it extremely hard to stay awake, even when the need to do so is great. Hormonal changes, for example in puberty, make certain behavior patterns highly likely, and in general, the neurochemical milieu has a powerful effect on the strength of desires, urges, drives, and feelings.

These considerations motivate thinking of control as coming in degrees, and hence as falling along a spectrum of possibilities. Toward opposite ends of the self-control spectrum are prototypical cases that contrast sufficiently in behavioral and internal features to provide a foundation for a basic, if somewhat rough-hewn, fuzzy-bordered distinction between being in control and not, between being responsible and not. In fact, as we consider various points on the spectrum, it seems likely that there are in fact *many* parameters relevant to being in control. Consequently, we should upgrade the simple one-dimensional notion of a *spectrum* to a multi-dimensional *parameter space*, where the dimensions of the parameter space reflect the primary determinants of in-control behavior.

In our current state of knowledge, we do not know how to specify all those parameters, nor how to weigh their significance. Nor are the relations between the parameters likely to be linear. We can, nevertheless, make a start. We do know now that activity patterns in certain brain structures, including the anterior cingulate cortex, hypothalamus, insula, and ventromedial frontal cortex are important. For example, large bilateral lesions to anterior cingulate abolish voluntary movement.⁹ One fortunate patient recovered some voluntary function. She also had good memories of her symptomatic period, during which, she explained "nothing mattered" and that she "had nothing to say."¹⁰ Smaller lesions to the anterior cingulate are associated with severe depression and anxiety¹¹ (Figure 1.1).

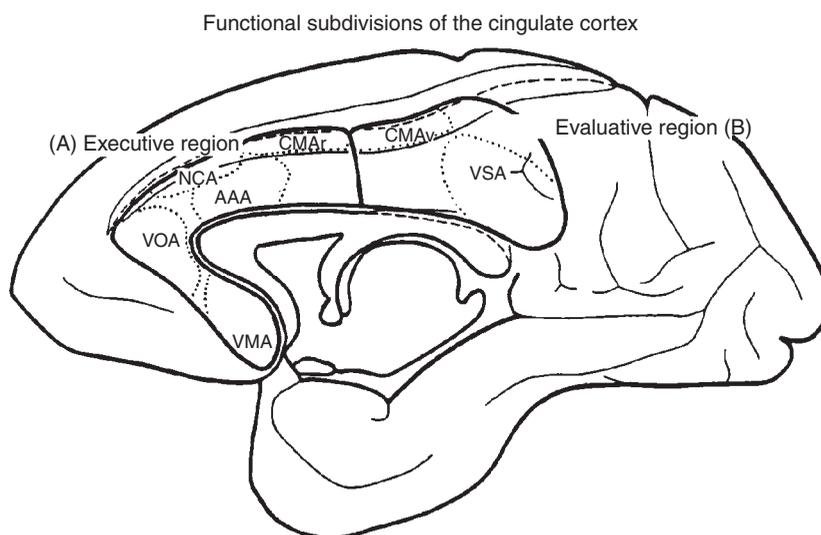


Figure 1.1 Functional division of the cingulate cortex of the rhesus monkey brain. The executive region (A) and the evaluative region (B) are the two major divisions. Subdivisions in (A): visceromotor (VMA); vocalization (VOA); nociceptive (pain) (NCA); rostral cingulate motor (CMAr); and attention to action (AAA). Subdivisions in (B) ventral cingulate motor (CMAv) and visuospatial (VSA) (based on Vogt, Finch, & Olson, 1992).

If the middle area of the cingulate is where the lesion occurs, patients may show loss of voluntary control over a hand. In *alien hand syndrome*, as this deficit is called, the hand behaves as though it has a will of its own. To the consternation of the patient, the hand may grab cookies, or behave in socially inappropriate ways. One patient discovered he could regain some control over his misbehaving alien hand if he yelled at it, “stop that!”

Imaging data implicate the anterior cingulate gyrus in the exercise of self-control over sexual arousal. In a functional magnetic imaging resonance (fMRI) study, male subjects were first exposed to erotic pictures, and then were asked to inhibit their feelings of sexual arousal. Comparisons between the two conditions show that when subjects are responding normally to erotic pictures, limbic areas show increased activation. When subjects engage in inhibition of sexual arousal, this activation disappears, and the right anterior cingulate gyrus and the superior frontal gyrus become more highly activated.¹²

The anterior cingulate again emerges as a player in autism. One undisputed finding is that autistics have deficits in analyzing affective signals. Because limbic structure play a central role in affect, a leading hypothesis claims the primary basis in autism to reside in defective affective evaluation, resulting from structural abnormalities in the limbic system.¹³ This hypothesis has been