

Prefrontal Cortical Deficit and (In)Compatibilism: Can Neuroscience Resolve the Free Will Problem?

There is a beguilingly simple (and familiar) argument that puts human freedom in doubt and with it nothing less than the foundations of morality. If moral agents do not act freely, then they are not responsible for their actions and, in that case, moral assessments of an agent's action have no principle. The case for some of these points may be efficiently stated since freedom of human action and moral responsibility seem to involve the same necessary condition: that an agent could have acted otherwise. The culprit is often cited follows:

- (1) If determinism is true, no one acts freely.
- (2) If no one acts freely, then no one is responsible for his/her actions.
- (3) Determinism is true. So,
- (4) No one acts freely and no one is responsible for his/her actions.

Determinism may be expressed in the following terms: for every event, e , there exists some other event, c , such that it is nomologically *impossible*, given c and the state of the universe at $t - 1$, together with the general and causal laws of nature, for e *not* to happen at t . On this view, human action is nomically necessitated. **Free** action, however, has involved two considerations: (i) even though an agent has performed a given action, she could have performed some other action instead and (ii) the agent originates her action or, at least, she has some significant measure of control over her action. Point (i) has emerged as a necessary condition on both freedom of action and moral responsibility. The convergence of freedom of action, moral responsibility and the power to do otherwise has a straightforward explanation; if moral responsibility requires freedom of action and freedom of action requires the ability to do otherwise, then moral responsibility

requires the ability to do otherwise.¹ In what follows, I explore the significance of connecting **responsibility** to an ability of the agent, specifically, her *ability to form a (special kind of) preference* for one action over another. More significantly, I will urge that the ability to form such preferences is not, considered in itself, causal; it is instead computational. I explain how, under certain conditions, such preferences causally necessitate *action*. Even so, an agent's ability to do otherwise is more a matter of what preferences are formed but since an agent controls her own deliberations, she *eo ipso* controls the computational processes that end with the formation of a preference and, in this way, she exercises her ability to do otherwise. It may be that progress on the free will question begins by rearranging the notions of free action, responsible action and acting otherwise. I begin with an account of a recent finding from neuropsychology for, I will contend, there is a useful notion of responsibility suggested therein.

I

Adrian Raine, et al., (2000) have produced striking evidence that there is a neurological abnormality underlying excessive and recurring violent behavior in certain individuals.² Such persons, brain imaging studies reveal, suffer an 11% reduction of gray matter nerve cell volume in the prefrontal cortex. Furthermore, the "prefrontal subjects" participating in Raine's brain imaging study were diagnosed (through widely recognized psychiatric standards as set forth in the *Diagnostic and Statistical Manual of Mental Disorders*) with a psychological condition known as anti-social personality disorder (APD). The disorder can involve a lack of control over rage that often ends in violence ranging from simple assault to homicide. Interestingly, it is possible to reliably

discriminate between control subjects and persons with a psychometric diagnosis of (APD) on the basis of brain imaging studies and subnormal autonomic responsiveness (autonomic responsiveness is often measured by galvanic skin response (GSR), (GSR) is the skin's resistance to conducting a low voltage electric current). What is more, although discrimination between groups, normal controls versus those in whom (APD) is present, may be reliably made on the basis of psychosocial risk factors (e.g., parental social class, early parental divorce, parental verbal arguments, parental criminality, parental physical fights and childhood abuse), the gain in discriminatory power is significantly advanced when autonomic responsiveness and prefrontal deficit is factored. The study is of special interest in that Raine has statistically demonstrated a relative independence between cortical deficit and the aforementioned psychosocial risk factors. As he puts it, "These analyses indicate that prefrontal and autonomic deficits in persons with (APD) cannot be attributed to psychosocial risk factors." Raine's study is also noteworthy in the effort made to control for comorbidity of (APD) with substance abuse and other psychiatric conditions. Patients in these populations do not present with cortical deficit.

It is not clear to what extent these individuals can be blamed for acts of violence; at least, it is not clear that they are *fully* responsible for such acts. Even so, these questions need not be considered, let alone settled, now. What is of interest is the possibility that prefrontal cortical deficit suggests another view of responsibility, primarily, and free action, secondarily. There are, it seems, important psychological and cognitive abilities that counterfactually, if not causally, depend on neurological structures that prefrontal subjects evidently do not have. Because such subjects

exhibit an underlying prefrontal deficit, they may exercise the associated abilities only within narrow limits, if at all. Acting nonviolently (or less violently) depends on certain psychological abilities, e.g., the ability to first delay an energized impulse to act with violence, the ability to act on foreseen consequences (say, undesirable outcomes of violence aimed at others or less undesirable outcomes in more restrained action) and, in light of such considerations, to *compute* the implications of such outcomes for preference formation and satisfaction. The computational and broadly logical processes that support such features of cognition clearly require an underlying neurology and insofar as impulse suppression is not degraded specifically (it seems), an intact prefrontal cortex. It is reasonable to conclude that a sustainable allegation of responsibility for violent action in those with prefrontal deficit would need to be measured against these points. Moreover, there is ample reason to believe that prefrontal subjects fail, I would say dismally, the two standard tests for free action: there is an identifiable set of conditions under which such persons cannot act otherwise and the manner in which their action is originated is not within their control. In what follows, I want to explore the possibility that Raine's prefrontal subjects are useful contrasting models for purposes of illuminating and reorienting our notions of free action, moral responsibility and an agent's ability to do otherwise.

II

Premise (1) of the argument given above has been skillfully advanced by philosophers such as Carl Ginet (1966, 1980), David Wiggins (1973) and Peter van Inwagen (1975, 1983) who have offered what is known as the "Consequence Argument" in its support.³ A summary version

of the Argument can be given with ease; if determinism is true, then no agent has the ability to do otherwise (and is, therefore, not free). One's action is jointly determined by past circumstances and the laws of nature, so, one could have acted in some other way only if either it is in one's power to undo or redo past circumstances or it is within one's power to make the relevant law of nature false (the relevant law(s) are those that necessitate the agent's action and, therefore, undercut an agent's ability to do otherwise). Since no one is able to undo or redo the past and no one is able to make the relevant law(s) of nature false, no one is able to do otherwise.

Classical compatibilists have made attempts to reconcile free action with its being caused.⁴ On some (historically prominent) versions of compatibilism, liberty is primarily a matter of acting as directed by one's will and without impediment or constraint. So, to say that an agent acts freely implies (i) she is able to act otherwise (in that her action is unimpeded) and (ii) she would have acted otherwise, if she had so willed or desired to act otherwise. Other songs in a similar key follow: (i') free action is a matter of doing otherwise and (ii') the meaning of 'doing otherwise' is expressed counterfactually; an agent would have done otherwise, if some contrary to fact condition(s) had obtained.

Modern versions of compatibilism have been located in one or another of two categories: reactive attitude theories and so called "mesh" theories. Broadly, mesh accounts contend that free action involves an (appropriate) mesh of the agent's actions with her reasons for acting. A widely cited example (of one kind of mesh theory) is that offered by Frankfurt (1971).⁵ Other types of mesh theory include those proposed by Gary Watson (1975) and Susan Wolf (1990).⁶ These

accounts postulate a link between, on one hand, an agent's system of values (as informed by the regulation of desire through reason) and free action (Watson) while, on the other, free actions are those which (essentially) involve "doing the right thing for the right reasons" (Wolf). Reactive attitude theorists (most notably P.F Strawson) place emphasis on the features of action that are required for an assessment of blame (or praise, as the case may be). In Strawson's view, responsibility is a function of two factors: (i) attitudes, for instance of resentment or gratitude, which may surface in response to the actions of others and more crucially (ii) the condition of being a "fit" object of such attitudes. Of course, the notion of "being fit" carries a considerable theoretical burden even if it is a good bet that this point has not bypassed Strawson.

In what follows, I consider, first, the relation of responsible action to its having been causally necessitated and, second, I develop a view, that I take to be essentially compatibilist, of how caused action is to be reconciled with free action, even if such action is nomically necessitated. The manner in which I reconcile responsible action with its being caused requires that events on the "input" or stimulus side of an agent cause or activate throughput *and* a caused action on the "output" or response side, subsequent to the end of throughput. The mechanisms constituting throughput need not involve anything that is essentially linked to (informed) values or "right reasons." Further, deterministic event causation can be propagated to a point well past input to include causal necessitation between certain mental events and overt behavior, the upshot is that the account I'll offer is consistent with the continuity of causal necessitation up to the point of deliberation as well as causal necessitation of action on the post throughput side.

III

Instead of taking acting otherwise (or the ability to do otherwise) as a necessary condition on **free** action, we might first consider the relation between **responsible** action and an **agent's ability** to make certain kinds of calculations (as against being **able** to **act** otherwise). The approach is to first develop a view of the conditions involved in responsibility and then consider free action in their light.

There are clues in the contrast between normal persons and the predicament of Raine's prefrontal subjects that suggest how the other approach might go. I begin by asking whether there is a difference between normal persons and individuals with prefrontal deficit that explains a neutral observer's hesitation to fully blame a person with prefrontal deficit for certain action while he or she might be considerably less hesitant in assigning blame to a normal person for the same action, under the same environmental conditions. The region of the brain which suffers a volume reduction in neural tissue is functionally tied to impulse suppression, as it seems. Normal persons, unlike prefrontal subjects, have the ability to suppress an energized impulse to act. This ability gives them enough time, and it seems not to require much, to consider future consequences of action. The impulse suppression circuits are "shorted" in prefrontal subjects, so they react instantaneously and often very violently (*even if they "see" a more advantageous choice*). What is more, their impulsivity does not allow for deliberation and they are unable to act on the obvious prediction that violence, under the circumstances, will lead to a term in jail (as in fact it has in

their past). A normal person is able to foresee, under similar environmental conditions, that violence will result in legal trouble and, on that basis, he or she will opt for restraint over violence.

The assessment that prefrontal subjects are not fully responsible for some of their actions is connected to a further fact, a fact that goes beyond the possibility of acting otherwise; for I have yet to consider the matter (or perhaps I should say, “manner”) of origination. In the case of such individuals there is little weight given to forethought in connection with some of their actions, they seem to react violently as if that is a default response. Moreover, that their reaction is a function of a congenital condition (in some cases) clearly raises questions surrounding the degree of blameworthiness in these persons. A charge of blame involving the action of a normal individual would clearly not be approached in the same way. The point can be developed through a fictitious example.

Smith has an enviable profession; he is a fishing boat Captain and three or four times each week he conducts deep sea charters on the Pacific coast, north of San Diego. The average charter consumes the better part of a day and Smith earns an often generous sum from between fifteen and twenty enthusiastic game fishermen. Smith is also an indefatigable card player with woeful luck. Loan sharks, and less unsavory creditors have their hooks in him and their patience with Smith’s excuses for not clearing his debts is now sorely tested. Smith has additional worries; as he looks out toward the horizon from shore, he sees some very ominous and very dark clouds. Being a fishing boat captain, Smith has specialized knowledge about weather patterns but he knows, like any competent adult and some children, what dark, ominous clouds portend. Even so,

he can't cancel the day's charter for he badly needs cash or the loan sharks may make good a certain threat made in connection with Smith's kneecaps. He sets out of the harbor, hoping the clouds will dissipate but catastrophe is not averted and twelve persons are drowned in a horrific storm. Smith is called before the Maritime Board to explain his action but the hearing does not go well. He cannot explain setting out to sea, ignoring the obvious threat of heavy weather. Of course, Smith cannot reveal certain other material facts, his need for cash and his bad luck at cards and, fortunately, the Maritime Board does not learn of the pertinent details. Even so, Smith has no reasonable explanation for his misdeed and that is bad enough. Smith is found blameworthy for not heeding a sign that any land dweller could understand. He is given a costly fine and his Captain's license suspended indefinitely.

I have a theory that describes and explains how an agent forms a (mere) **preference** for his actual action rather than another. Let's begin with the supposition that agents are subject to two types of (reason-directed) desires, approach (reason-directed) desires and avoidance (reason-directed) desires. I want to allow for the possibility that desires can conflict; so there are approach-approach desires, approach-avoidance desires and, lastly, avoidance-avoidance desires. Now let us reconsider the tragic tale of Jones. On the day of the tragedy Jones might have thought better of putting out to sea; he could have canceled that day's charter and, had he done so, so many lives would not have been consigned to a watery grave. Of course, if he had done so, he would have to reckon with the loan sharks and other creditors. Smith has a choice between two actions and he has a reason-directed desire to do either one. The suggestion is that certain desires, unlike those

that surface in homeostasis, are shaped by a cognitive component, nicely captured with the phrase, “reason-directed desire.” Jones is now pulled in two directions by conflicting reason-directed desires. He cannot follow both courses of action; he must settle for one or the other. It is the need for resolution that “activates” or causes a computational state; as in, “I could cancel the charter and, perhaps, stall for time with the loan sharks. But I’ve been stalling for weeks, and besides, the sharks probably have hired help to carry out their threats, after all, they can’t garnish wages or ruin one’s credit rating; they must have alternative measures for collecting debts. I cannot risk ignoring their threats but I can risk bad weather. If the clouds are this close, I’ll know within an hour of setting sail if the weather is going bad and we won’t be that far offshore.”

The last two points in Jones’s deliberations are critical for they lead to a preference that is crucially related to his subsequent action; Jones now **prefers** to go on with the charter. So his current mental state is such that he has a preference for going forward with the charter rather than the only alternative action of cancelation. However, with some obvious change in details, he could have formed a preference for canceling the charter. Such details, as we shall see, create the possibility of acting otherwise. Even so, the point I want to stress now is this: the computation results in the transformation of what was before a conflict between two reason-directed desires, a reason-directed desire to err for the sake of caution and cancel the charter and a reason-directed desire to avoid an outcome connected with a threat and put to sea, to a preference concurrent with resolving that conflict. The effect of the newly formed preference dampens the reason-directed desire to err on the side of caution and cancel the charter and, consequently, Jones’s prior state of

indecision is effectively over. Preferences of this kind are to be distinguished from both a reason-directed desire and certain other kinds of desire states, . The idea is that the agent moves from a reason-directed desire to act in one way and a reason-directed desire to act in another way to a preference for acting in one way rather than the other.

I want to introduce a distinction between a *mere* preference versus a *potentiated* one. The distinction can be explained by noting that although it is obvious what action or series of actions would result in the satisfaction of Jones's (mere) preference, there are further questions for Jones to consider. In sum, these questions revolve around the foreseeable consequences of alternative actions open to Jones ("foreseeable" by a standard of what may be reasonably expected from a person of average ability, mental and physical, and common knowledge). Such questions may or may not occur to Jones. But, on the hypothesis that they do, Jones would thereby be caused to engage **another** computational state. That state is, as before broadly logical, but it is also distinctive in that Jones's mere preference is transformed into a potentiated one. Exactly how this happens involves a bit of speculative psychology but the idea is plausible. The transformation of a mere preference into a potentiated preference is effected, in part, by Jones's additional finding, revealed in another computation, that among the consequences of canceling the charter, there is no *compelling* reason to do so even if there is a good reason to have done so (some care is required on this point; Jones may not judge that there is a compelling reason in the consequences of alternative open actions to refrain from acting on a preference and he may or may not judge that there is good reason to refrain from acting on a preference in the same set of consequences).

Jones's mere preference becomes potentiated by virtue of his conjoining that preference with his *judgment* that there is no compelling reason to refrain from acting on the preference furnished in the foreseeable consequences of the only open alternative open action. Remember, he thinks, *wrongly* as it sadly turned out, he can deal with the weather's going bad and so he is prepared to assume that risk. One may well suppose that under these circumstances Jones is primed for action; he has a potentiated preference *and* an ill-considered (and mistaken) judgment that there is no compelling reason to refrain from acting on the preference. He has now moved past a threshold and what he will end up doing is as predictable as any human action can be.

On the account developed so far, I have insisted that the inferential, computational states involved in elevating a mere preference to a potentiated preference is not, considered in itself, causal. Such states can be caused and they end with potentiated preferences that cause actions but there are two reasons for thinking that the relevant computational states are not causal. First, there are reason-directed desires, which are unlike desires that regulate physiological homeostasis, in that such desires have propositions for their associated complement sentence or "that clause." Certain other kinds of desires, involved in regulating homeostasis often have a direct object in place of a proposition. Mere and potentiated preferences also have propositions as complement sentences. This is not accidental, propositions have truth values. And it may be that assessments of blame depend upon an impartial observer's finding that a crucial preference, of either kind, or a reason-directed desire was *incorrectly* accepted as true by the agent being assessed. Such reason-directed desires and preferences are a *species* of propositional attitude and the only kinds of

relations *these* attitudes can have to one another are inferential, rather than causal. Second, an agent's computational states are not regulated by laws of nature and, most certainly, such states are not regulated by the formal principles of sound reasoning; clearly, agents often make mistakes of reasoning and such errors may be just as relevant to assessments of blame as facts of common knowledge. It is entirely likely that standards for assessment will depend upon some reasonable notion of what can be expected from a mentally competent adult of average intellectual ability and common knowledge. Since the relevant kinds of attitudes, preferences, and computations are not regulated by the laws of nature, the relevant inferential, computational states are not, considered in themselves, causal, although, as I have urged, potentiated preferences, in virtue of certain complex causal properties, can and do cause action.

IV

The following question may now be faced: what light does the preceding shed on responsibility, an agent's ability to do otherwise and, ultimately, free action? Jones's mere preference is transformed into a potentiated one as the story unfolds, but it has, while causing him to do what he ends up doing, now trumped the only other competing reason-directed desire. It is thusly transformed by none other than Jones himself insofar as he makes a second determination that, among the consequences of the only alternative open action, there is no compelling reason to refrain from acting on the (mere) preference of doing the charter. Upon transformation of a mere preference into a potentiated preference, action subsequent to and in the service of a potentiated preference is thereby causally necessitated. At that point, the formerly open alternative to act

otherwise is effectively closed. However, if neither preference becomes potentiated, and until one or the other mere preference is transformed into a potentiated preference by Jones's computational effort, either action is open for Jones and he would have acted in some other way, if his deliberations had been different with respect to the relevant details. Even so, an allegation of blame is supported by the fact that the transformation (of a competing mere preference) into a potentiated preference is effected by Jones's *other* computations. In substance, the point is familiarly made in terms of origination. Alternatively, an agent originates her action in that she brings about (through some quasi-logical or any rate non-causal computational state) the potentiation of a competing mere preference because it is *she* who determines that among an alternative action's consequences there is no compelling reason to refrain from acting on the preference.

There are many occasions in which it is reasonable to hold a normal person, who has the relevant skills, responsible for her action. But there are also cases of caused action for which a person is not fully responsible and Raine's prefrontal subjects make that point well enough. This is consistent with the claim that, under ordinary conditions, a normal person can be held responsible for her action, even if the action is causally necessitated by a potentiated preference and especially if along side the manner of origination there are features, for example, a mistaken but avoidable, determination that no alternative outcome provides a compelling reason to refrain from acting on the relevant preference. So, some responsible actions are caused actions. Such action is event caused by a potentiated preference for that action over another but a normal agent

is responsible for having acted all the same since she may determine, rightly or wrongly, but in either case through her own computations that other outcomes furnish no compelling reason to have refrained from so acting.

What matters to responsible action is origination by virtue of an agent's ability to form, through an inferential, computational state, a potentiated preference. Such preferences then cause a specific action. Remember, mere preferences are formed from a reason-directed desire, and so, mere preferences have some causal properties since desires have causal properties. When a competing mere preference is potentiated, it acquires additional causal properties since the potentiated preference effectively "suppresses" all other competing reason-directed desires and any associated mere preferences; a preference is potentiated by the agent's additional finding that none of the consequences of alternative open actions furnished a compelling reason to refrain from acting on what will soon become the one and only potentiated preference. What's needed to get to "free will" compatibilism is the premise that caused action is compatible with responsible action. The latter point is the missing piece in the puzzle: what's *required* for a finding of blame is that an agent would have formed a potentiated preference for some *other* action, if her deliberations had taken a different course. And, if her deliberations had been different, she would have acted in some other way. The implication is that the prior and necessary condition for being blameworthy may not so crucially depend on an ability to act otherwise, since a potentiated preference necessitates the agent's action and, if anything, it now "suppresses" any inclination to act otherwise. Moreover, since preference potentiation is computational, there will be, in the

typical case, certain events that cause or activate the inferential, computational state that underlies, but does not cause, the transformation of a competing mere preference into a potentiated preference. But since a potentiated preference is a cause of responsible action, compatibilism emerges from a straightforward argument: (1) any responsible action is a free action. (2) Some responsible actions are caused actions. Consequently, (3) some free actions are caused.

Finally, there may be a way out from underneath the Consequence Argument. Perhaps there is a third possibility upon which an agent's ability to act in some other way depends. As it turns out, unlike past circumstances or the laws of nature, an agent exercises complete control with respect to the third hypothesis. As we have seen, unless and until a mere preference is transformed into a potentiated preference, any of a number of different actions, each with an associated reason-directed desire and a competing mere preference if one is formed, are open possibilities for an agent. If an agent's deliberations had been different, some other mere preference would have become potentiated and the agent would have acted in some other way. And since the agent has full control over the course of her deliberations, she controls what mere preference is transformed into a potentiated preference and, to this extent, she has control over alternative actions as well as her ability to act in some other way.

¹ The claim that moral responsibility requires the power to do otherwise is contentious. Harry Frankfurt challenged the point in his, "Alternate Possibilities and Moral Responsibilities," Gary Watson, ed., 2nd edition, *Free Will* (Oxford: Oxford University Press, 2003), p. 169.

²Raine, A., et. al., 2000. "Reduced Prefrontal Gray Matter Volume and Reduced Autonomic Activity in Antisocial Personality Disorder." *Archives of General Psychiatry* 57:119-127. Other neural structures may be contributory; see Raine, A. et. al. 2003. "Corpus Callosum Abnormalities in Psychopathic Antisocial Individuals." There are studies that tend to confirm Raine's initial (2000) findings, e.g., van Honk, Jack et. al. 2001. "Repetitive Transcranial Magnetic Stimulation at the Frontopolar Cortex Reduces Skin Conductance but not Heart Rate: Reduced Gray Matter Excitability in Orbitofrontal Regions/In Reply" (by Raine), *Archives of General Psychiatry*, Vol. 58. 10: 973-975. More importantly, other confounding, comorbid conditions have been suggested (some of which seem to have been well controlled in Raine's 2000 study). The abnormality may have a perinatal correlation (it is possible, for example, that smoking during pregnancy is implicated) see Raine, A. et. al. 2005 "Neurocognitive Impairments in Boys on the Life-Course Persistent Antisocial Path" *Journal of Abnormal Psychology*, Vol. 114 (1) Feb 2005: 38-49. The upshot: underlying organicity in prefrontal subjects appears to play a significant role in subverting their ability to act responsibly under certain conditions.

³Ginet, Carl. "Might We Have No Choice." in Lehrer, Keith (ed.) 1966. *Freedom and Determinism*. New York: Random House: 87-104, "The Conditional Analysis of Freedom." in van Inwagen, Peter. (ed.) 1980. *Time and Cause*. Dordrecht: Riedel: 171-186; Wiggins, David. "Towards a Reasonable Libertarianism." in Honderich (ed.) 1973. *Essays on Freedom and Action*. London: Routledge & Kegan Paul: 31-61; van Inwagen, Peter. 1975. "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27: 189-199, 1983. *An Essay on Free Will*. Oxford: Clarendon Press. The argument, as formulated by van Inwagen, has been critically examined by others; see the essay by Tomis Kapitan in Kane, Robert. 2002. *The Oxford Handbook of Free Will*. Oxford: Oxford University Press and, especially, Lewis, David. 1981. "Are We Free to Break the Laws?" *Theoria* 47: 112-121.

⁴The term 'classical compatibilism' was introduced by Watson, Gary. 1975. *Journal of Philosophy* 72 (8): 205-20, reprinted in Watson, G. 2003. *Free Will*. 2nd edition. Oxford: Oxford University Press. In the classical view, free action is unconstrained ability to do as one wills.

⁵Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68: 5-20, reprinted in Watson (2003).

⁶Watson, Gary. 1975. "Free Agency." *Journal of Philosophy* 72: 205-220. Reprinted in Watson (2003). Wolf, Susan. 1990. *Freedom within Reason*. Oxford: Oxford University Press.