PRINCIPIA **29**(4): 707–713 (2025) doi: 10.5007/1808-1711.2025.e107648
Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

NOTAS E DISCUSSÕES
NOTES AND DISCUSSIONS

RESPONSIBILITY PROFILES: A NOTE ON ADRIAN RAINE

MARCELO FISCHBORN
Farroupilha Federal Institute of Education, Science, and Technology, BRAZIL
marcelofischborn@gmail.com
https://orcid.org/0000-0002-0410-5251

Abstract. At least since the 1970s, there has been an expectation that results from cognitive neuroscience will eventually impact the practices associated with the concept of moral responsibility. Those practices include expressions of blame and the institution of legal punishment. The standard approach to those changes appeals to skepticism about free will and moral responsibility. However, that approach faces a major challenge: the persistence of beliefs and attitudes about free will, moral responsibility, and the appropriateness of punishing wrongdoers. This paper seeks to develop an alternative approach, based on the concept of a responsibility profile. Empirical work, such as Adrian Raine's, has shown that individuals vary in how they react to being held responsible. Those variations may help to inform the proposal and assessment of better ways of implementing the practices associated with moral responsibility in ways that do not require skepticism.

Keywords: moral responsibility • responsibility profile • punishment • free will

RECEIVED: 25/06/2025 ACCEPTED: 28/10/2025

At least since the pioneering work by Benjamin Libet in the 1970s (Libet et al. 1983), there has been an expectation that cognitive neuroscience and other sciences of the mind will at some point impact questions related to free will, moral responsibility, and the practices associated with those concepts. A standard approach to the reform of those practices begins by claiming that a given set of empirical findings justifies skepticism about the reality of free will and moral responsibility and concludes that the institution of legal punishment and prevailing ways of expressing blame should be revised or abolished. That standard approach has been pursued in discussions of the empirical findings of Libet and others (see, for example, Greene & Cohen 2004; Libet 1999; Sapolsky 2024; Sinnott-Armstrong & Nadel 2011). The standard approach

© 2025 The author(s). Open access under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License.

is often followed by a standard pushback: scholars, especially philosophers, have replied that the empirical findings under discussion are either insignificant or insufficient to undermine free will and moral responsibility (see, for example, Fischer 2023; Mele 2009).

In *The Anatomy of Violence* (2013), Adrian Raine follows the standard approach. His work focuses on violent behavior, which has greater practical significance than Libet's finger movements in the lab. As a psychologist working on the social, biological, and neural causes of antisocial behavior, Raine provides evidence supporting the hypothesis that the brains of violent offenders, including serial killers, differ from more average ones. On the practical implications of his findings, he says he believes that "in tomorrow's world we can rise above our feelings of retribution, reach out for rehabilitation, and engage in a more humane discourse on the causes of violence" (Raine 2013, p.372). More specifically, he argues that violent people with brain abnormalities caused by disadvantaged genetics, gestation, nutrition, or upbringing have less freedom to control their behavior and therefore less moral responsibility. As a result, they deserve less severe punishment for the crimes they may end up committing.

Instead of presenting a standard pushback, I will discuss some aspects of Raine's work that do not need to be framed as the standard attack on free will and moral responsibility. Some of his proposals leave room for an alternative approach to questions about how responsibility practices should be implemented. While the standard approach starts by examining the causal factors that influence decision-making and behavior, there is an approach to be developed by emphasizing the causal factors that impact *reactions to being held responsible*. After examining Raine's views in more detail, I elaborate on that alternative approach by proposing the concept of a responsibility profile and elaborating on its place in empirical investigations about how people vary in their reactions to being held responsible.

Raine reviews a broad literature on biological and social factors whose interaction can predispose someone to violence. According to him, there is a "recipe" for violence, whose ingredients include factors related to genetics, gestation, nutrition, abuse, and hormones, all of which can leave a mark in the brain. Raine summarizes those findings in a model that links brain-related risk factors to cognitive, affective, and motor predispositions to violent behavior. The alternative approach I want to propose begins by making a distinction among the items present in Raine's model. Here is a lengthy quote that will help to frame the discussion:

> Impairment [to brain areas such as the ventromedial prefrontal cortex and the medial-polar prefrontal regions] results in poor planning and organization, impaired attention, the inability to shift response strategies, poor cognitive appraisal of emotion, poor decision-making, impaired self-reflection, and *reduced capacity to adequately process rewards and punishments.* These cognitive impairments translate into social elements that lead to crime—

> poor occupational and social functioning, noncompliance with societal rules, *insensitivity to punishment cues that guide behavior*, bad life decisions, poor cognitive control over aggressive thoughts and feelings, overreaction to minor irritations, lack of insight, and school failure. [...] Impairment to [regions such as the amygdala, the hippocampal complex, and the insula] can result in an inability to understand the mental states of others, learning and memory impairments, lack of disgust, impaired moral decision-making, *lack of guilt and embarrassment*, lack of empathy, poor fear conditioning, poor emotion regulation, and *reduction in uncomfortable emotions associated with moral transgressions*. [...] Brain impairments [in areas such as the dorsolateral prefrontal cortex and the orbitofrontal cortex] result in response perseveration, motor impairments involving a failure to inhibit inappropriate responses, impulsivity, the *failure to shift response sets and passively avoid punishment,* [...] poor impulse control, *the failure to avoid punishment* and disruptive behavior. (Raine 2013, p.267–269, emphasis added)

On a conceptual level, the items included in this model are not equally relevant to questions about moral responsibility. The standard approach emphasizes factors affecting decision-making and action. They matter for attributions of free will and moral responsibility. Items in Raine's model, such as "poor planning", "poor decision-making", and "poor cognitive control", fit nicely within that approach. Such dysfunctions invite us to assess whether freedom and responsibility are undermined. A second group of items, in contrast, includes "reduced capacity to adequately process rewards and punishments", "insensitivity to punishment cues", "lack of guilt and embarrassment", "reduction in uncomfortable emotions associated with moral transgressions", and "failure to avoid punishment". The dysfunctions of this second group concern how people respond to being held responsible. While the first group is mostly concerned with what happens before an action is performed, the second group focuses on what happens thereafter.

The processes of the first group concern someone's decision-making profile. Raine's findings suggest that different people may follow different patterns, or be biased in different ways, by the processes that influence their choices and actions. The processes of the second group relate to what I would like to call a *responsibility profile*: a particular way someone reacts (or even fails to react) to being held responsible, actually or potentially. For example, variations in the processes of the second group can make different people respond differently to expressions of blame, reprehension, or punishment by others. Those variations can also make different people vary in how (if at all) they self-direct responses, such as guilt, embarrassment, and other negative emotions usually associated with moral violations. On the positive side, similar variations can also affect how people respond to praises or rewards. Acknowledging such variations is just a first step, to be followed by the classification and identification of specific types of responsibility profiles.

Of course, processes related to decision-making and responsibility profiles interact and may be difficult to distinguish in practice. For instance, does a chronic reoffender have impaired decision-making or failure to avoid punishment? It may be both, yet emphasizing the failure to avoid punishment may favor a different approach to how a person should be held responsible, both within the legal sphere and in more ordinary contexts.

Alternative approaches to reforming responsibility practices should be welcomed because the standard approach faces significant challenges. One of them, which proponents of the standard approach have acknowledged (Greene & Cohen 2004, p.1783–1784; Sapolsky 2024, Chapter 14), is the robustness of ordinary beliefs about free will and moral responsibility and the desire to see wrongdoers punished (see also Bigenwald & Chambon 2019; Fischborn 2018, 2024). One plausible explanation is that attitudes about moral responsibility and punishment evolved as part of the mechanisms that enabled human cooperation (Fischborn 2023; Greene & Cohen 2004, p.1784; Mameli 2013; Sapolsky 2024). If that is true, the disposition to hold people responsible has deep roots in human moral psychology, which make it unlikely for human societies to abandon practices that seek to blame and punish violations of significant norms. But robustness does not equal immutability. Responsibility practices, punitive ones in particular, have been changing throughout human history (Morris & Rothman 1995; Sapolsky 2024, Chapter 14) and even those who believe in the reality of free will and moral responsibility can agree that current practices may be defective and in need of further change (Fischborn 2024; Mele 2013, p. 189; Strawson 1962, p.170).

Although defining the concept of a responsibility profile is a philosophical task, specifying which responsibility profiles exist requires empirical investigation. There are some initial candidates worth exploring. For example, Edmund Rolls (2000), whose work is cited by Raine, investigated how the orbitofrontal cortex influences responses to punishments and rewards. In monkeys, damage to that brain area impairs the learning of associations between stimuli and rewards, as well as changes in those associations. That damage impairs performance in tasks that require learning about "which stimuli are rewarding and which are not, and especially in altering behavior when reinforcement contingencies change" (Rolls 2000, p.285). Humans with damage to the frontal lobes often present similar difficulties: "frontal patients may be able to verbalize the correct rules [required to succeed in an experimental task] yet be unable to correct their behavioral sets or strategies appropriately" (p.289) and they may also have difficulties with gambling tasks and persist in choices that are "no-longer-rewarded" (p. 290). Those kinds of cognitive impairments, in turn, can cause problems in daily tasks and social relations. For example, some patients manifest "euphoria, irresponsibility, lack of affect and lack of concern for the present or future" (p.289). Caregivers also report "disinhibited or socially inappropriate be-

havior; misinterpretation of other people's moods; impulsiveness; unconcern or underestimation of the seriousness of their condition; and lack of initiative" (p. 290). These findings suggest a potential variable to be included in a responsibility profile. There are individual differences in how well people make associations between their choices and positive or negative outcomes, as well as in how they adjust subsequent behavior based on those associations. In other words, some people have deficits in their capacity to learn the negative consequences of their actions and to behave accordingly.

Another, and related, potential variable to be included in a responsibility profile is what Philip Jean-Richard-Dit-Bressel, Simon Killcross, and Gavan P. McNally (2018, p.1645) refer to as "punishment sensitivity". They note that many branches of neuroscience study the mechanisms involved in processing punishments. One object of study is decision-making in general, but other studies focus on individual disorders. As they say, "sensitivity to punishment is assessed across a variety of disorders, including addiction, depression, psychopathy as well as eating disorders, enabling insights into the etiology, maintenance, and treatment of these conditions" (p.1639). They list substance use disorders, gambling disorder, obsessive-compulsive disorder, antisocial personality disorder, conduct disorder, and oppositional defiant disorder as examples that have been linked to reduced sensitivity to punishments (p.1645). Depression, in contrast, has been linked to increased sensitivity to punishment (p.1646). Although many details about punishment processing and potential new treatments are still missing, the relevant observation now is that something in line with the concept of a responsibility profile is already present in the investigation.

A major challenge for the standard approach to reforming responsibility practices is the robustness of belief in free will, moral responsibility, and the appropriateness of delivering punishments to wrongdoers. One reason to think the approach centered on responsibility profiles is not similarly affected is the finding that decisions about how to punish are sensitive to considerations about their consequences. Empirical evidence supports the claim that people's decisions about how to punish someone are influenced by what they expect will result from different punitive options (Fischborn 2024). If new treatments are developed that target disorders associated with specific responsibility profiles, it seems a plausible hypothesis that information about the effectiveness of such treatments can also influence decisions about how to punish and how to hold responsible more generally.

An additional challenge for initiatives seeking to reform responsibility practices is that punitive desires tend to increase in the face of immoral or criminal behavior (Clark et al. 2014; see also Spitzley 2021). If alternatives are to be embraced by society and implemented, we will need robust evidence that they can do a better job at preventing offenses and recidivism; otherwise, they may end up backfiring. That is a challenge for both the standard approach and the alternative focusing on

responsibility profiles. The responsibility profiles alternative, however, may be more promising in this regard because it starts with the assumption that individuals will behave differently when they are held responsible. It is conceptually tied to the idea that different ways of holding people responsible may have different effects.

The standard approach has dominated the literature. Generations of neuroscience studies have been interpreted mainly as challenges to free will and moral responsibility. Alternative conceptual frameworks may have a better chance of translating the relevant empirical findings into improved social practices, as was argued above. Works by Raine (2013), Rolls (2000), and Jean-Richard-Dit-Bressel, Killcross, and McNally (2018), among many others, help to point to new possibilities. Those possibilities may benefit from a new approach—especially one that is not centered on issues about free will and moral responsibility—and the concept of responsibility profile may help in this regard.

References

Bigenwald, A.; Chambon, V. 2019. Criminal Responsibility and Neuroscience: No Revolution Yet. *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.01406

Clark, C. J.; Luguri, J. B.; Ditto, P. H.; Knobe, J.; Shariff, A. F.; Baumeister, R. F. 2014. Free to Punish: A Motivated Account of Free Will Belief. *Journal of Personality and Social Psychology* 106(4): 501–513.

Fischborn, M. 2018. How Should Free Will Skeptics Pursue Legal Change?. *Neuroethics* 11(1): 47–54. https://doi.org/10.1007/s12152-017-9333-8

Fischborn, M. 2023. The Evolutionary Roots of Moral Responsibility. *Philosophy of Science* 90(4): 817–835. https://doi.org/10.1017/psa.2023.60

Fischborn, M. 2024. Reforming responsibility practices without skepticism. *Philosophical Psychology* 37(4): 904–920. https://doi.org/10.1080/09515089.2022.2090913

Fischer, J. M. 2023. Review of *Review of Determined: A Science of Life Without Free Will*, by R. M. Sapolsky. *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/reviews/determined-a-science-of-life-without-free-will/

Greene, J.; Cohen, J. 2004. For the Law, Neuroscience Changes Nothing and Everything. *Philosophical Transactions of the Royal Society of London* 359: 1775–1785.

Jean-Richard-Dit-Bressel, P., Killcross, S., & McNally, G. P. 2018. Behavioral and neurobiological mechanisms of punishment: Implications for psychiatric disorders. *Neuropsychopharmacology* 43(8): 1639–1650. https://doi.org/10.1038/s41386-018-0047-3

Libet, B. 1999. Do We Have Free Will?. *Journal of Consciousness Studies* 6(8–9): 47–57.

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(3): 623–642.

Mameli, M. 2013. Meat made us moral: A hypothesis on the nature and evolution of moral judgment. *Biology & Philosophy* 28(6): 903–931. https://doi.org/10.1007/s10539-013-9401-3

Mele, A. 2009. *Effective Intentions: The Power of Conscious Will*. New York: Oxford University Press.

Mele, A. 2013. Free Will, Science, and Punishment. In: T.A. Nadelhoffer (ed.), *The Future of Punishment*, p.177–191. New York: Oxford University Press.

Morris, N.; Rothman, D. J. 1995. Introduction. In: N. Morris & D. J. Rothman (eds.), *The Oxford History of the Prison*, p.vii–xiv. New York: Oxford University Press.

Raine, A. 2013. *The Anatomy of Violence: The Biological Roots of Crime*. New York: Pantheon Books.

Rolls, E. T. 2000. The orbitofrontal cortex and reward. *Cerebral Cortex* 10(3): 284–294. https://doi.org/10.1093/cercor/10.3.284

Sapolsky, R. M. 2024. *Determined: A Science of Life without Free Will*. Penguin Press.

Sinnott-Armstrong, W.; Nadel, L. (eds.). 2011. *Conscious Will and Responsibility: A Tribute to Benjamin Libet*. New York: Oxford University Press.

Spitzley, J. 2021. The Future of Moral Responsibility and Desert. *Review of Philosophy and Psychology*, *12*(4): 977–997. https://doi.org/10.1007/s13164-020-00522-5

Strawson, P. F. 2009 [1962]. Freedom and Resentment. In: D. Pereboom (ed.), *Free Will*, p.148–171. Indianapolis: Hackett.

Acknowledgments

I wrote this paper as a postdoctoral researcher at the Federal University of Rio Grande do Sul. I thank Rogério Severo for feedback and for supervision, and the Farroupilha Federal Institute of Education, Science, and Technology for the research leave.