2025, 11 (1)

ARGUMENTA

The Journal of the Italian Society for Analytic Philosophy

First published 2025 by the University of Sassari

Produced and designed for digital publication by the *Argumenta* Staff

Argumenta is the official journal of the Italian Society for Analytic Philosophy (SIFA). It was founded in 2014 in response to a common demand for the creation of an Italian journal explicitly devoted to the publication of high quality research in analytic philosophy. From the beginning *Argumenta* was conceived as an international journal, and has benefitted from the cooperation of some of the most distinguished Italian and non-Italian scholars in all areas of philosophy.

Contents

Editorial

This is the last editorial to bear only my signature.

After ten years of life, and in reaffirming its identity as a Diamond Open Access journal, *Argumenta* is undergoing a profound restyling, beginning with the Editor-in-Chief role. From now on, this position will be jointly held by Cristina Amoretti (University of Genoa), Filippo Ferrari (University of Bologna), and myself (University of Sassari). Needless to say, I am extremely excited to be part of this inspiring trio. And this is not the only good news.

The entire Editorial Board has been substantially expanded. On the one hand, Section Editors and Junior Section Editors have been appointed for each philosophical area covered by the journal. On the other hand, a group of Managing Editors, supported by young Assistant Editors, has been established. The Book Review Editors have been confirmed and will continue to play the invaluable role they have demonstrated so far. I warmly invite you to visit the Argumenta website to find out more about the new editorial team and the journal's latest developments.

As usual, I will now briefly introduce the present issue.

It opens with a Topical Collection that, to my knowledge, has few equals in the current literature: *The History and Philosophy of Free Will*, edited by Sofia Bonicalzi and Mario De Caro. The collection ranges from ancient theories of action—such as those of Plato and Aristotle—to Moritz Schlick's radical critique of traditional debates on free will, passing through critical analyses of the arguments of Spinoza, Leibniz, Locke, Hume, Kant, and Hegel. Although free will is an evergreen topic whose appeal shows no sign of waning among scholars

Argumenta 11, 1 (2025): 3-4
ISSN 2465-2334

and non-specialists alike—and although many important insights will undoubtedly be offered in the years to come—I am deeply convinced that this Topical Collection represents an ineluctable milestone in this centuries-old debate.

This issue also includes three articles that have already appeared in Early View, by Juha Räikkä, Marco Hausmann, and Sanjit Chakraborty. They have already made, and will continue to make, significant contributions to discussion in their respective fields.

The issue concludes with the Book Reviews section. We are pleased to offer our readers three thoughtful reviews of three noteworthy recent books.

Finally, I would like to thank all the colleagues who served as external referees, the members of the Editorial Board, the editors of the Topical Collection, the Book Review editors, and the team of librarians at the University of Sassari—especially Francesca Madrau and Maria Caterina Arru. All of them have been extremely generous with their time, work, advice, and suggestions.

As usual, the articles appearing in *Argumenta* are freely accessible and freely downloadable, therefore it only remains to wish you:

Buona lettura!

Massimo Dell'Utri
Editor-in-Chief

Argumenta 11, 1 (2025)
Topical Collection

The History and Philosophy of Free Will

Edited by

Sofia Bonicalzi and Mario De Caro

The Journal of the Italian Society for Analytic Philosophy

Contents

The History and Philosophy of Free Will: An Introduction

Sofia Bonicalzi and Mario De Caro

Roma Tre University

Free will has long occupied a central place in philosophical thought, tracing a rich trajectory from the reflections of the ancients to the systematic treatments of the contemporary age. Far from being a purely historical concern, it continues to command attention in current debates, especially within the analytic tradition, where the tension between alternative accounts of freedom sparks ongoing philosophical discussion.

Long before compatibilism and incompatibilism became philosophical jargon, the debate was already divided between those who—anticipating what would later be called compatibilists—argued that free will could coexist with determinism. They maintained that even if all events were determined by prior causes, individuals could still be free and morally responsible for their actions, provided they acted according to their desires, reasons, and intentions without external coercion. In contrast, their intellectual counterparts, the incompatibilists, held that determinism and genuine freedom were fundamentally incompatible. Across the centuries, these basic positions have persisted in various forms. Among the views that remain most well-known are libertarianism, which holds that agents have indeterministic freedom and the genuine capacity for alternative possibilities; hard determinism, which denies the existence of free will entirely, asserting that all events, including human choices, are causally determined; and mysterians, who argue that the nature of free will is ultimately an insoluble mystery, beyond human understanding.

These enduring distinctions carry significant implications for ethics, responsibility, and our understanding of human agency. If determinism is true, can praise, blame, or moral accountability be justified? If indeterministic freedom exists, how is it exercised, and what are its limits? And if free will is ultimately mysterious, what does that imply about our attempts to resolve these questions?

Tracing the roots of the contemporary debate, this special issue aims to elucidate the contributions of significant figures in the history of philosophy, from the ancients through the early twentieth century, while simultaneously addressing theoretical questions of enduring relevance. Our goal is not merely to document historical developments but to engage critically with the conceptual challenges that continue to animate contemporary discourse. This special issue thus brings together contributions from a range of scholars, including several leading experts, offering original insights and advancing the ongoing dialogue on free will. As evidenced by the vitality of the contemporary debate and the richness of these

Argumenta 11, 1 (2025): 9–12
ISSN 2465-2334

DOI 10.14275/2465-2334/202521.bon
First published: 30 November 2025

papers, far from being a historical relic, the study of free will remains a vibrant and contested field, illuminating the complex nature of human agency and the conditions under which individuals can be held morally accountable.

In *From Guilt to Responsibility: Ancient Theories of Action from Homer to Aristotle (plus Alexander of Aphrodisias)*, Carlo Natali explores how early discussions of responsibility, initially raised by poets and legislators, gradually evolved into systematic philosophical inquiry. He shows how these early reflections on human action and accountability laid the foundation for a more rigorous, conceptual treatment of moral responsibility. Plato sought to challenge the authority of the poets through a rational, Socratic-inspired approach, emphasizing reasoned reflection over mythic narrative. Aristotle, building on and refining this tradition, developed a comprehensive account of human action that evaluates the full spectrum of deeds, rather than focusing solely on wrongdoing or blameworthy acts. Natali emphasizes that Aristotle's notion of responsibility is grounded in a precise ontology of events, linking moral psychology to natural philosophy and highlighting the complex interplay between choice, character, and circumstance. Extending this line of thought, Alexander of Aphrodisias argued that, in any given situation, an agent could choose either of two possible actions—a position supported by the Aristotelian conception of causation as a one-way dependence that does not entail strict necessity. Overall, Natali's study demonstrates both the sophistication and the subtlety of ancient approaches to human responsibility, revealing a nuanced understanding of choice, agency, and moral accountability that laid the groundwork for subsequent philosophical inquiry throughout the Middle Ages and into the early modern period.

Over the centuries following antiquity, thinkers such as Augustine, Aquinas, Erasmus, and Montaigne, among others, transformed and enriched debates about the will, moral responsibility, and human agency. This long trajectory of reflection set the stage for early modern thinkers to critically reassess the very possibility of freedom, culminating in the radical critiques of the seventeenth century. In particular, Spinoza's rejection of free will in the *Ethics* has long been understood as a challenge to the very idea that the will can be genuinely free. In his *Spinoza on Freedom, Feeling Free, and Acting for the Good*, Leonardo Moauro revisits this critique, offering a refined reinterpretation of its scope and implications. While Spinoza not only denies the freedom of the will but also explains the widespread belief in it as the product of our limited self-knowledge, traditional readings tend to focus primarily on his arguments against incompatibilist conceptions of freedom. Moauro contends that this focus is too narrow, showing that Spinoza's position also undermines compatibilist accounts, according to which freedom consists in self-determination guided by value judgments. On this view, human actions are not directed by rational conceptions of the good but are driven by appetites indifferent to such evaluative ideas, while the sense of acting freely arises from projecting value onto the objects of desire. By denying that we possess the capacity to determine our actions in accordance with our conception of the good, Spinoza rejects a feature commonly regarded as central to agency, thereby revealing his critique of free will to be far more radical and philosophically far-reaching than is often assumed.

Leibniz, in turn, is often regarded as occupying an uneasy position in debates on free will, given his early and persistent skepticism regarding the very coherence of the notion. Throughout his philosophical career, he treats free will not only as dispensable for grounding moral responsibility but also as a potential obstacle to a proper understanding of divine justice and human retribution. In *Free Will in*

Leibniz's Thought, Gianfranco Mormino addresses this tension by reconsidering the role that contingency plays in Leibniz's account of free action. While Leibniz famously insists that free actions must be contingent, Mormino argues that the ability to have acted otherwise, understood in terms of alternative possible worlds, does not, in itself, constitute freedom. Rather, Leibniz conceives freedom as a form of perfection rooted in rational activity, in which actions align with reason and the agent's internal principles. Contingency, on this view, serves not to ground freedom but to explain the possibility of error, showing how human actions can go astray without presupposing an indeterministic will. By disentangling freedom from modal alternatives, Mormino offers a clearer and more nuanced picture of Leibniz's conception of agency and responsibility, demonstrating how moral accountability can be preserved even in the absence of libertarian free will.

The special issue then includes two contributions that examine the relevant work of British empiricists. In *Locke on Free Will and Epistemic Responsibility*, Samuel C. Rickless offers a nuanced analysis of Locke's conception of freedom, framing it as the mind's capacity to act in accordance with its own volitions. For Locke, freedom is a power that inheres only in substances, and humans exercise it in a variety of contexts, including the capacity to suspend the immediate pursuit of their desires. This suspension, however, is not morally neutral; it represents a critical moment for reflection, during which individuals are expected to deliberate on their options and determine which actions are most likely to promote their overall happiness and well-being. Rickless highlights a central tension in Locke's account: human deliberation is susceptible to motivated irrationality, meaning that agents may be led astray by self-interest, bias, or emotion. To act freely and responsibly, individuals must not only weigh their choices but also continually monitor the rationality of their own reasoning. This recursive demand—reflecting on one's own capacity to reason—risks an infinite regress, presenting a profound philosophical challenge and raising enduring questions about the scope and limits of moral responsibility, as well as the feasibility of consistently exercising rational self-governance.

Remaining within the same geographical and intellectual context, Hume's treatment of free will is inseparable from his broader project of developing a science of human nature grounded in experience, sentiment, and social practice. In *Hume on Free Will*, Lorenzo Greco reconstructs this account by drawing on Hume's discussions in both the *Treatise of Human Nature* and the *Enquiry Concerning Human Understanding*, showing how his compatibilism gains its full significance within an empiricist, naturalistic, and skeptical framework. From this perspective, human beings are understood as reasonable and active agents whose experience is structured by causal relations that render both individual action and social interaction intelligible. Greco argues that, within such a framework, the problem of free will does not turn on abstract metaphysical concerns about determinism, but on the practical conditions under which responsibility is attributed. For Hume, responsibility is rooted in sentiments of approval and disapproval directed at agents, insofar as their actions express stable character traits, dispositions, and patterns of conduct that are empirically observable and causally explicable. Greco concludes by emphasizing that, although free will may present difficulties in moral or religious contexts, these challenges do not undermine Hume's account, which remains firmly anchored in human practices rather than metaphysical speculation.

The problem of reconciling human freedom with a deterministic understanding of nature is a central challenge for a Hume admirer like Kant. In *Kant on Free Will*, Derk Pereboom examines how Immanuel Kant confronts this tension,

carefully distinguishing his account from Humean compatibilism, Spinozist determinism, and various indeterministic approaches. Kant's theory is both ambitious and nuanced: he defends a form of agent-causal libertarian free will, which he terms transcendental freedom, while simultaneously affirming that the natural world operates under strict causal determinacy according to established scientific principles. At the same time, he is careful to acknowledge the limits of theoretical demonstration; transcendental freedom cannot be established empirically, and our evidence cannot confirm it as a fact of the natural world. Instead, Kant grounds our belief in freedom on practical reasons, arguing that recognizing our capacity for self-determined action is necessary for coherent moral deliberation and the justified attribution of responsibility. By articulating this careful balance, Kant's account preserves the possibility of genuine moral agency without undermining the causal order of nature, offering a distinctive and deeply considered solution to one of the most enduring tensions in philosophical thought.

Remaining within the German philosophical tradition, Thomas Meyer, in *Hegel on Free Will*, offers an insightful and carefully argued analysis of Hegel's conception of freedom, a notion that is central not only to his practical philosophy but also intimately connected with his philosophy of mind, his philosophy of nature, and the *Science of Logic*, forming a cornerstone of his entire philosophical system. Meyer begins by tracing the philosophical motivations and historical influences that shaped Hegel's system, before presenting the account of free will introduced in the *Elements of the Philosophy of Right*. Although this account leaves some space for skepticism, Meyer goes further, reconstructing the underlying metaphysical framework of Hegel's notion of freedom as developed in the *Science of Logic*. Through this reconstruction, Meyer illuminates Hegel's effort to navigate the tension between human freedom and the causal determination of the world. He concludes by articulating what he calls Hegelian libertarianism, an incompatibilist and indeterminist interpretation of free will that underscores both the depth and the remarkable complexity of Hegel's thought.

Finally, turning to the early twentieth century, in *Free Will: A Pseudo-Problem? Schlick on a Longstanding Metaphysical and Ethical Debate*, Sofia Bonicalzi examines Moritz Schlick's radical critique of traditional debates over free will, situating his intervention within the framework of logical empiricism. Schlick, a leading figure of the Vienna Circle, argues that much of the controversy arises from persistent conceptual and linguistic confusions surrounding terms such as "freedom," "will," and "determinism." Bonicalzi shows how Schlick's neopositivist approach, marked by a rejection of metaphysical speculation and abstract normative theorizing, effectively reframes the problem: once these concepts are clarified, the apparent paradoxes of free will largely dissolve. At the same time, Schlick acknowledges the continued practical relevance of moral and legal responsibility, provided these issues are addressed within a clear empirical and causal framework. By situating Schlick's critique within the broader philosophical and scientific context of his era, Bonicalzi highlights both the originality and the limits of his position, presenting free will as a pseudo-problem generated by theoretical ambiguity rather than a substantive metaphysical or ethical dilemma.

Taken together, these studies highlight the enduring philosophical engagement with human freedom, the nature of action, and moral responsibility, showing how thinkers across history have explored the capacities, limits, and conditions that define human agency.

From Guilt to Responsibility: Ancient Theories of Action from Homer to Aristotle (plus Alexander of Aphrodisias)

Carlo Natali

Ca' Foscari University of Venice

Abstract

In Greece, the problem of responsibility for choices was originally discussed by poets and legislators. Philosophical analysis developed in relation to these reflections, and Plato wanted to undermine the authority of poets with a new approach, inspired by Socrates. Aristotle tackled the question starting from Plato's position. His approach was not limited to the question of the responsibility of those who perform evil actions, but extends to a general evaluation of the entire sphere of human actions. The problem of responsibility in Aristotle is discussed on the basis of a precise ontology of events, and falls within the field of moral psychology and physics. Alexander of Aphrodisias even came to support the thesis that in two identical situations we always have the possibility of choosing one action or its opposite. This is based on the very complex Aristotelian idea of causation as a one-way dependence relationship, which does not always imply necessity.

Keywords: Aristotle, Poetry, Responsibility, Causality, Determinism.

1. Some Preliminary Notions

The philosophical discussion on guilt and responsibility in ancient Greece develops against a particular cultural background, which must be taken into account for the historical understanding of the terms of the debate and to avoid a too hasty identification of ancient positions with contemporary problems. In the beginning, the problem of responsibility for choices was not a philosophical issue, but was discussed by poets, legislators and tragedians; philosophical analysis developed in relation to these reflections, sometimes in continuity with them, more often in opposition to the opinion of poets and literati.

The great epic poems as the *Iliad* and the *Odyssey* were not seen as mere works of art and a source of aesthetic pleasure, in the manner of today, but were considered, from the Archaic age until the Christian age, a guide on how to live.

Argumenta 11, 1 (2025): 13–33
ISSN 2465-2334
DOI 10.14275/2465-2334/202521.nat
First published: 30 November 2025

Plato, who disliked this aspect of the culture of his time, testifies that "the praisers of Homer [...] say that this poet educated Greece" (*Republic* X, 606e1-3).[1]

The world of epics is peopled by men and gods, two races similar in appearance and character; the gods are immortal, infinitely more powerful than men, and are endowed with a peculiar clarity of vision, while a mist obscures human eyes. In spite of an asserted contempt for men, the gods love to intervene in their world, directing their actions, but the divine action on earth is discreet, almost subterranean, hardly ever evident or prodigious. In order to influence the actions of men, sometimes the gods present themselves in human form and give interested advice to those they address; sometimes, however, the gods operate within the individual, and in some way possess him. They can thus enhance the warrior's fury (menos), or obscure his ability to decide (atē).

Thus one of the most important problems in the Homeric epics arises, that of the moral responsibility for evil actions. Some of the most important actions of kings and heroes, such as the kidnapping of Helen by Paris, were inspired by a god or a goddess, and one may wonder whether such actions fall under the responsibility of those who perform them. In fact the Homeric heroes sometimes do, with a full personal participation, what a god imposes them to do, and accept to pay the penalty of the faults committed by external impulse. Agamemnon and Helen, in famous passages, blame themselves for their actions, whatever the gods have done. For example Agamemnon says:

> 1. However I am not at fault (aitios), but Zeus and Moira and Eryns that walks in darkness […] they cast upon my soul atē on that day […] but what I could do? It is god that brings all things to their issue […] seeing that I was blinded, and Zeus robbed me of my wits, fain I am to make amends and pay liberal gifts (*Iliad* XIX, 86-90 and 137-8; transl. Murray [Homer 1924] modified).

The atē reinforces the condition of blindness typical of the human condition, and according to these passages one can be guilty of an act without being its actual cause. Modern critics have debated these passages endlessly, and the prevailing opinion is that in the *Iliad* human will and divine influence are totally integrated, unreflectively, with a kind of double motivation. These words constituted a problem for philosophical reflection as early as the first century A.D., when Plutarch wrote that:

> 2. People despise Homer and say that with his impossible exploits and incredible tales he makes it impossible to believe in every man's rational power to make a choice (ton hekastou logismon tēs proaireseos apiston kathistanos) (*Life of Coriolanus* 32; transl. Perrin [Plutarch 1916] modified).

Plutarch thus acknowledges that poetic reflection on causality and responsibility for actions was important in the debate in ancient Greece.

In poetry, however, we should not look for an iron coherence; in the *Odyssey* Zeus seems to answer to Agamemnon and polemicize with the idea that the gods make men commit faults:

[1] All translations from the works of Plato are taken from Plato 1997, with occasional modifications.

3. It is strange how ready mortals are to blame the gods and say we inflict the evils they bear, but they even by themselves, through they blind folly, bring sorrows on themselves beyond the will of fate (*Odyssey* I 32-34; transl. Murray [Homer 1919] modified).

The *Odyssey* departs from the archaic perspective of the *Iliad*, and in this poem another aspect of action comes to the fore, the idea that the agent acts of his own volition (hekōn), when he performs an action. For example Menelaus at one point says, "Not of my own will (hekōn) do I stand still" (*Odyssey* IV, 377 [Homer 1919]).

The Homeric poems thus already provide a good deal of the typical terms of the debate on determinism ancient thought: aitia, aitios, hekōn etc. One thing is missing: will. We cannot here investigate the question in all its aspects, but we would like to note that in Greek culture the idea of a psychic faculty such as the will, autonomous and independent both of reasoning and of emotions and desires, has not been present for a long time.

Another source of reflection on culpability was the legal lexicon, in which it was important to distinguish between crimes committed ex pronoias, premeditatedly, and those akōn, unintentionally, as the Law of Drakon of the fifth century BC puts it. An ancient inscription says:

4. If a man without premeditation kills (me 'k pronoias kteinei) another, he will be brought to trial. The Basileis are to judge the accused person […], the Ephetai are to decide the case […] if he killed unintentionally (akon) and the fifty-one Ephetai recognise that he killed unintentionally (akonta), let ten phratry members readmit him to the homeland (*Inscriptiones Graecae* I3 104, lines 11 and 16-18; I follow the interpretation by Pepe 2012 and 2019).

Pronoia implies forethought, epiboulē kai proboulē, as the orator Antiphon in a trial speech attests:

5. I will prove that my opponent's mother murdered our father in a planned and premeditated way (ex proboulēs kai proboulēs) […] he has been the involuntary (akousiōs) victim of a planned murder (ex proboulēs), she sent him to death deliberately and with premeditation (hekousiōs ex pronoias) (*Against the stepmother*, §§ 3 and 5).

This distinction raises the problem of how to evaluate actions performed neither out of premeditation nor involuntarily, but out of anger and immediate impulse (thumōi, dia thumon). In this period they were often considered akōn, not voluntary, in the sense of not premeditated (ouk ethelōn, akōn, *Iliad* XXIII, 87-8 [Homer 1924] and Euripides, *Herakles* 1363-4).

The specifically philosophical reflection at its beginning starts precisely from investigating the responsibility of the wicked for their actions. Socrates' thesis, reported by Plato, says that:

6. I am pretty sure that none of the wise men thinks that any human being willingly (hekonta) makes a mistake or willingly (hekonta) does anything wrong or bad (*Protagoras* 345e1-4).

In short, Socrates holds that everyone in their actions tends to their good but errs in identifying it, and that moral guilt consists in a mistaken identification of the good with the pleasure or profit of the moment. The question of re sponsibility for error is reduced to an inability to calculate correctly the human good in a given situation. His thesis is thus that ethical virtue is identified, at bottom, with a form of knowledge (*Protagoras* 355a3-b3). Socrates refers to divine intervention in a very limited way, and in him the question of responsibility arises on the human level only.

2. Plato

Plato's position is rather clear when viewed against this background. Broadly speaking, it can be said that one of the philosopher's intentions was to undermine the authority of Homer and ancient tragedy as the "masters of Hellas", because he is convinced that they convey to their audience bad conceptions of divinity, virtue, and responsibility for actions. This is a critique that is at once political, theological, and moral, and in fact is found in Plato's most political dialogues, such as the *Republic* and the *Laws.* In these dialogues Plato designs societies with a high degree of perfection, and refuses to allow both epic poetry and tragic plays within his ideal cities. Nevertheless, Plato used the vocabulary of poetry and tragedy to formulate his theses.

In a very brief summary, Plato on the one hand accepts the thesis that no one does evil voluntarily:

> 7. When men whose physical constitutions are bad have bad forms of government where bad civic speeches are given, both in public and in private […] that is how all of us who are bad come to be that way, being the product of two causes completely involuntary (dia duo akousiōtata) (*Timaeus* 87a7-b4).

And:

> 8. All wicked men are, in all their mistakes, unwillingly (akontes) wicked (*Laws* IX, 860d1).

But at the same time Plato rejects the Homeric thesis that the god is responsible for the evils that the individual attracts to himself by his own evil actions. The god is good and therefore:

> 9. What is good isn't the cause of all things, but only of good ones; it isn't the cause of bad ones […] Therefore, since the god is good, he is not—as most people claim—cause of everything that happens to human being but only of a few things […] we must find some other cause for the bad ones, not the god (*Republic* II, 379b13-c7).

In the myth that concludes the *Republic,* everyone chooses the kind of life they want to live, and they are not forced by god:

> 10. The responsibility lies with the one who makes the choice: the god has none (aitia helomenous: ho theos anaitios) (*Republic* X, 617e4-5).

From our point of view, Plato's position can be characterized as a separation between voluntariness and responsibility: if one misidentifies his own good, he commits evil unwillingly (akōn); but he is equally responsible (aitios) for what he does, even if he is influenced by a corrupt environment or psychophysical defects. In a way, paradoxically, Plato recovers an archaic position, with the only difference, very important for him, that the gods are exonerated from any responsibility for the crimes committed by men.

As for actions performed by impulse, Plato considers them an intermediate case (metaxu*, Laws* IX, 867a1), and not a clear one: there are impulsive actions that are closer to the involuntary, those performed immediately, others closer to the voluntary, those performed after a certain lapse of time (*Laws* IX, 866d-867a). The problem seems evidently to consist in the fact that between 'premeditated' and 'against will' there is an intermediate space, which does not fit into either category.

Apparently, in Plato's school, the Academy, it was customary to discuss the master's theses with great freedom, and the disciples felt free to reject some of Plato's theses. Most likely there was also a debate about the question of responsibility. We have some traces of it in Plato himself and, after him, in Aristotle in particular.

In Plato's *Laws* the position is referred of someone who tried to distinguish between the character of the individual and his individual actions. This person argued that individual action is performed voluntarily, even though the individual character that determines it is involuntary:

> 11. To suppose that a voluntary act is performed involuntarily makes no sense. Therefore, in the eyes of someone who holds the view that injustice is involuntary, a man who acts unjustly would seem to be doing so against his will […] If anyone with a disputatious disposition or a desire to attract favourable notice says that, although there are those are unjust against their will (akontas), even so many men do commit unjust acts voluntarily (hekontas), I reject their argument and stick to what I said (*Laws* IX, 860d9-e4).

Plato rejects this thesis:

> 12. Here and now, that is the position I have to accept: I allow that no one acts unjustly except against his will (*Laws* IX, 860d 8-9).

The most comprehensive discussion, however, is to be found in the ethical works of Aristotle, who devotes two large sections of the *Eudemian Ethics* (II, 6-10) and the *Nicomachean Ethics* (III, 1-5) to discussing this question, with some major innovations.

3. Aristotle

Aristotle's position is not only linked to the question of the responsibility of those who perform evil actions, but extends to a general evaluation of the entire sphere of human actions, good and bad, on the basis of a precise ontology of events. With this move, the discussion on determinism takes a decisive step, leaving a perspective originally limited to the juridical and moral field to invade also the sphere of physics and metaphysics. Aristotle introduced into the debate some key terms, which would be taken up by the following philosophical

schools for centuries after him. Aristotle does not approach the problem of determinism in terms of fate (heimarmenē), although this term had already made its appearance in Plato and in his successor as head of the Platonic school, Speusippus, who devoted a specific treatise to this subject (*Peri heimarmenēs*, fr. 43 Heinze). He prefers to discuss the status of contingent events and the statements concerning them.

4. A Thought Experiment

Let us begin with a logical work, *De interpretatione*, Chpt. 9. This is a celebrated text, in which Aristotle, after stating the general dialectical principle that for every contradiction one of the members of the contradictory pair is true and the other false (*De int.* 6), lists a number of exceptions to this rule, including sentences concerning contingent future events such as "Tomorrow there will be a naval battle". Aristotle affirms that, if every statement of this kind is true or false, then everything necessarily happens or does not happen (18a34-5). Obviously, as he notes in the treatise entitled *Categories*, it is not the truth of the proposition what the necessity of the event depends on, but, on the contrary, only if a future event happens necessarily, the sentence in the future tense describing would be true (*Categ.* 4a35-b2, cf. *De int*. 18b37-9). Because of that, the true Aristotelian thesis is that every sentence in the future tense would be true or false if and only if everything were necessary. Indeed, by admitting that every proposition about future contingents is true or false, something absurd derives:

> 13. If it is white now it was true to say earlier that it would be white; so that it was always true to say of anything that has happened that it would be so. But if it was always true to say that it was so, or would be so, it is not possible not be so, or not be going to be so. And if something cannot not happen it is impossible for it not to happen; and if it is impossible for something not to happen it is necessary for it to happen. Everything that will be, therefore, happens necessarily, and nothing will come about as chance has it or by chance; for if by chance, not of necessity (*De int*. 18b9-16).[2]

There is a kind of thought experiment in this piece, which in simplified form, might sound like this:

> 14. (1) If at time *t* (present), it is true to say, in a future tense sentence, that event x will occur at time *t*+1, then
> (2) at time *t*-1 it would have been true to say, in a future tense, that the event *x'*, (which actually occurred at time *t*), would have happened, and then it would have done so necessarily.

If things were so, it follows that nothing would happen by chance (tuchēi) or as it happens (hopoter'etuchen), for necessity and chance are opposed (18b5-7). But if nothing happens contingently, then:

[2] All translations from the works of Aristotle are taken from Aristotle 1984, with occasional modifications.

15. Everything is and happens of necessity. So there would be no need to deliberate or to take trouble (oute bouleuesthai oute pragmateuestai), thinking that if we do this, this will happen, but if we do not, it will not (*De int*. 18b30-3).

A counterpart of this passage is found in the *Metaphysics,* book VI (E), in a section devoted to the study of accidental events. Aristotle argues that, if of every event there is a non-accidental cause, then everything turns out to be necessary, and he confirms this with the same thought experiment. Compared to the previous passage, here there is no longer any reference to logic and language, but the idea is expressed in terms of the causal connection of events:

16. Will this be or not?—Yes if this happens; and if not, not. And this will happen if something else does. And thus if time is constantly subtracted from a limited extent of time, one will obviously come to the present [follows an example constituted by a human action] .. And similarly if one jumps over to the past, the same account will hold good; for this—I mean the past condition—is already present in something. Everything, therefore, that is to be, will be of necessity, e.g. it is necessary that he who lives shall one day die; for already something has happened (*Metaph*. 1027a32-b10).

In my opinion this passage illustrates the case that Aristotle wants to deny: it shows a chain of successive events, each of which is the cause of the other, a chain that is run backwards from the future to the present, when a certain event either happens or it does not. And from the present to the past, at the cause of such an event. But if every cause and principle is generated by a previous event, then everything is necessary. Here Aristotle repeats the same thought experiment as in *De interpretatione:*

17. (1) If at time *t* (present) we have an event x that causes in a proper sense an outcome *r* at time *t*+1.
(2) then at time *t*-1 an event *x′* had occurred which was a cause in the proper sense of the occurrence of event *x* at time *t*, which in fact did occur, and did so necessarily.

Aristotle wants to deny this consequence. In *De interpretatione* he affirms:

18. But this is impossible, for we see (1) that what will be has an origin both in our deliberation and in our acting (horōmen hoti estin archē tōn esomenōn apo tou bouleuesthai kai apo tou praxai ti), and that, (2) in general, in things that are not always actual there is the possibility of being and of not being; here both possibilities are open, both being and not being, and consequently, both coming to be and not coming to be. Many things are obviously like this. […] Clearly, therefore, not everything is or happens of necessity: (2.1) some things happen as chance has it (hopoter'etuche), and of the affirmation and the negation neither is true rather than the other; (2.2) with other things it is one condition rather than the other and for the most part (mallon [...] kai hōs epi to polu), but still it is possible for the other condition to happen instead (*De int*. 19a7-22).

This passage is usually understood as a polemic against so-called logical determinism, but in fact Aristotle starts from ontological considerations, and the distinction of three kinds of entities and events:

(1) human actions that are the result of deliberation and whose principle (archē) is in us,
(2) contingent events, divided into two sub-categories,
(2.1) things that happen indifferently one way and the other, and
(2.2) things that preferably happen a certain way, and rarely the opposite.

This second division could also apply to the category of human actions, but Aristotle does not seem to want to do so. He does not seem committed to demonstrating the existence of the contingent, but on the contrary starts from the existence of the contingent, given as evident, to illustrate the conditions of its occurrence, that is, that there must be a first cause such that it does not go back to anything else. For not everything to be necessary, one must arrive at a cause which is not generated by anything else. In *Metaphysics* this is an accidental cause:

> 19. Clearly then the process goes back to a certain principle (archē), but this no longer points to something further (ouketi eis allo). This then will be the principle of the fortuitous (tou hopoter' etuchen), and will have nothing else as cause/responsible (aitios) of its coming to be (*Metaph*. 1027b10-6).

This cause is called the "principle of what happens by chance" (1027b12-13).

Now, in *De interpretatione* and the *Metaphysics,* Aristotle mentions two kinds of causes or events not generated by anything else, (1) human choice and (2) the accident, but he never clarifies what relationship there is between these two kinds of cause. Critics have struggled for centuries to identify a relationship between the two spheres. It seems clear that an accidental event is not identical with an event that is the result of a single individual's choice, and that it does not exhaust the whole sphere of the contingent; in my opinion the most promising perspective is the one that makes the presence of contingent events a necessary, but not sufficient, condition of human choice.

Aristotle's position on human responsibility in *De interpretatione*, Chpt. 9, as we have seen. is based on this idea: man is archē of his own actions. Let's examine this idea in detail.

5. From Fault to Responsibility: *Eudemian Ethics*

In passages 16 and 18 Aristotle uses human actions as examples of non-necessary events: from every action of ours other events depend, and if our actions were necessary, all the events that depend on them would be necessary. Human action is seen by him as the first element of a chain: if we do *x, y* will happen, and he seems to be especially interested in stating that given a chain of events arranged in series, the beginning of the chain must not be determined by anything else on pain of determinism (*De int.* 18b14-5). In *Eudemian Ethics,* II 6, Aristotle reiterates his point, insisting forcefully on the idea that human action is part of the contingent and that he who is *archē* of an action (as text 18 says) is *aitios* of it*,* and vice versa. In this chapter a new term is introduced, 'what depends on us', indicating what we are the first cause of:

> 20. So that it is clear that all the acts of which man is the principle and master (archē kai kurios) may either happen or not happen, and that their happening or not happening—those at least of whose existence or non-existence he is master

> (kurios)—depends on him (eph'autōi). So he is the cause/responsible (aitios) of what it depends on him (eph'autōi) to do or not to do, and what he is the cause/responsible (aitios) of depends on him (eph'autōi) (*Eudemian Ethics* II,1223a1-9).

The final sentence indicates a biunivocal correspondence between what we are the cause of *x* and "what depends on us" *y*: *x*↔*y*. The term *archē* must not be interpreted neutrally, as a simple "starting point"; here it designates the "principle" of a thing, in the sense of that which has "authority, power" over it, in fact it is associated with the term kurios, "master".

In the *Eudemian Ethics* Aristotle explains that when man is the principle, archē, of the actions he performs (1222b19-20) this principle is to be understood in the sense of "that from which" movement derives. In this sense archē means an efficient cause (1222b30). Being the efficient cause of contingent events, this archē too is contingent:

> 21. Since some realities may have the opposite to their actual qualities (enia tōn ontōn enentioōs echein), so of necessity must be their principles (archai). For what results from the necessary is necessary; but the things said before can be different and many of them depend from men themselves, and men are the principles (archai) of such things (*Eudemian Ethics* 1222b41-1223a4).

The interpretation of this passage is controversial, but it seems to me to say that since human actions are among the contingent events, also the cause of human actions is contingent, in the sense that man can head once towards one choice and once towards the opposite one.

To express the idea that it depends on us whether our actions occur or not, Aristotle choices a phrase taken from everyday language, especially military language. In passage 20 the expression eph'hēmin (in the form eph'autōi), which was almost completely absent from the previous theoretical debate, makes its appearance. It can have two main meanings.

(1) In a local, proper, or figurative sense, it indicates that which is on our side, comes towards us, or against us, or in general affects us. We have examples of this usage in many authors of the fifth and fourth centuries, such as Thucydides, Euripides, Isocrates, Sophocles, Iseus, Demosthenes.
(2) In the "potestative" sense, however, it indicates what depends on us, as opposed to what depends on others, that is, what is in our power, especially in the case of war: e.g. victory, a prisoner, the enemy, booty. In this second sense it expresses the fact that if an action x depends on someone, non-x also depends on him. Examples of this sense are found in Sophocles, Xenophon, Hippocrates, Demosthenes, Alcidamas, and in the so-called Rhetoric to Alexander.

In this second sense the phrase implies something more than the simple concept of 'spontaneous', as an act performed without coercion, or the simple idea that the agent has various options to choose from. It indicates that the choice is up to him alone: examples of this strong usage are given in literature as well, e.g. in Euripides: 'the outcome depends on me (telos d'eph'hēmin)' (*Helen*, v. 878-893) and in an oration of Iseus: "when it depended on us (ot' eph'hēmin) to avenge ourselves [...] we did not want to (ouk eboulēthēmen)" (*De Dikaiogenes*, § 30, 5).

Aristotle with this lexical choice he seems to introduce a very strong condition for establishing what is voluntary, the fact that we are masters of our choices and we are the efficient cause of them. This condition overturns the legal distinctions of his time, for on the basis of it impulsive actions are transferred from the realm of the akon to that of the ekousion, of what is done voluntarily, since the efficient cause is in us (1111a20-25 + 1111b1-3). It solves in this way Plato's hesitations about how to categorise acts by thumos we saw before.

Aristotle in the *Ethics* nominalizes the expression with the neuter plural article: ta eph'hēmin, the things that depend on us. They are everything whose archē, the efficient cause, is in us, such as the objects of choices, deliberations, wills, relationships with friends, actions, moral qualities such as virtue and vice and in general habitual states, certain diseases of the body resulting from our disordered life, the value of money etc., while the rest, such as natural qualities, does not depend on us. The eph'hemin being of these things is to be understood distributively, in the sense that each of them individually depends on us.

Some have argued that for Aristotle we are not masters of the *choices* we make, but only of the *actions* we take, because he does not list choices among the things that depend on us. According to these critics, choices depend on character, and once a certain character is acquired, the choice becomes necessary, even if the action has us as its efficient cause. Some then argue that according to Aristotle what depends on us does so only at the generic level, as a *type of* action, but not at the specific level: even if the *type z* of actions depends on us, e.g. tying our shoes, it does not follow from this that doing so depends on individual *Y* at time *t*. Given his character *c* and situation *s,* his choice is already determined, and, according to them, *Y* does not have a dual path to take in front of him, but will necessarily tie his shoes.

I shall return to this question when speaking of character and habit, but I may say at once that the argument that our choices do not depend on us at first sight seems strange. In the current usage of poets and orators we saw that choices could be qualified as things that depend on us. In Aristotle, in particular, given that our actions result from our choices, if the choices we make are not dependent on us it is not clear how, within Aristotle's philosophy, our actions could be dependent on us.

In the *Nicomachean Ethics* he explicitly says that choice "is a voluntary thing (hekousion)" (1112a14). Sometimes Aristotle does not simply say that man is archē of his action, but more precisely that man's choice is the archē of actions. In Book VI he analyses choice by means of a distinction between efficient cause ("that from which") and final cause ("that in view of which"), saying:

> 22. Choice is the principle (archē) of action—"choice" in the sense of "that from which the movement derives" and not "that in view of which"—and desire and reasoning with a view to an end are the principles of choice (*Nicomachean Ethics* VI,1139a30-3).

In the *Metaphysics,* then, he confirms that "what is done and what is chosen are the same thing" (1025b24).

In passage 20 we saw that what depends on man, *ta eph'autōi,* is what he is responsible for, aitios. We have already seen that in Plato the individual is aitios, responsible for what he does, even if he performs it akōn, involuntarily. In

Aristotle, on the contrary, it is not possible for one to be aitios of a thing accomplished akōn, involuntarily:

> 23. We all admit that of acts that are voluntary and done from the choice (ekousia kai kata proairesin) of each man he is the cause/responsible (aitios), but of involuntary acts (akousia) he is not himself the cause/responsible (aitios); and all that he does from choice (proelomenos) he clearly does voluntarily (hekōn) (*Eudemian Ethics* II,1223a16-9).

Thus choice is hekousion, we are aitioi of it and, if we are aitioi, it is eph'hemin, it depends on us: we cannot be responsible for acts, even mental acts, that do not depend on us.

6. The Discussion in the *Nicomachean Ethics*

The *Nicomachean Ethics* is probably a treatise addressed to a fairly large audience, and is more discursive in tone than the *Eudemian Ethics*. In the section on "what depends on us" on book III, we see traces of Academic discussions (cf. § 2), to which Aristotle makes various references, though without naming specific opponents.

In this work to define what is voluntary, hekousion, he starts from the involuntary, akousion, which he divides into two parts: that which is done by force and that which is done through ignorance of the concrete conditions of action (1110a1); from which it follows that there are two necessary conditions for voluntary action:

> 24. The voluntary (hekousion) would seem to be (1) that of which the principle (archē) is in the agent himself, (2) he being aware of the particular circumstances of the action (1111a22-4).

The first condition is already known to us from passage 22: it is the fact that the first efficient cause must be internal to the agent; but the second condition is added only in this work. We are now mainly interested in the first condition. In the first chapter of *NE* III Aristotle establishes three important points. First of all, condition (1) states that actions performed by force (biāi), meaning those whose efficient cause is *totally* external, are not voluntary:

> 25. Those things, then, are thought involuntary, which take place under compulsion (biāi) or owing to ignorance; and that is compulsory of which the principle is outside (hē archē exōthen), being a principle in which nothing is contributed (mēden sumballetai) by the person who acts or is acted upon, e.g. if he were to be carried somewhere by a wind, or by men who had him in their power (1110a1-4).

Further on he specifies better that by the term archē he means the efficient cause: the "principle (archē) that moves the organic parts of the body" (1110a15-6).

The problem of action in the *Nicomachean Ethics* is set up through a strong insistence on the notion of causality. It is therefore important to clarify the characteristics of the Aristotelian notion of cause. For Aristotle "cause (aitia)" is a term that has various meanings, and generally indicates a one-way dependence of one entity or event on another ($a \rightarrow b$), which can take various forms: efficient, formal, final or material dependence. All of them can be organised in

form of a chain. Each type of causal chain indicates a single kind of dependence, and each one is finite, meaning that it depends on a first cause of the same type. Moreover, it is not possible to conceive of composite causal chains, which mix together different forms of dependence (*Metaphysics* II 2). And the relations between different causal chains are complex: not always a causal chain necessitates another. Indeed, for example, the sum of an efficient causation and a final causation does not give a necessary causal chain.

Since the notion of cause is multiple, it is possible to say at the same time that no event is without cause, but not all events are necessary. Even if it is true that a certain agent performs an action for a certain end, the final cause is not itself necessitating:

> 26. But if some one were to say (phaiē) that pleasant and noble objects (ta hēdea kai ta kala) have a compelling power, because they force us and are from without (anankazein gar exō onta), all acts would be for him compulsory; for it is for these objects that all men do everything they do (1110b9-11).

Here the contrast is between the case of text 25, in which the action is considered involuntary because there is an efficient external cause, to which the responsibility for the action must be attributed (the wind, the enforcers), while in this text 26 the presence of a cause which is external indeed, but is a final one, makes the action not necessary and does not make it involuntary.

Passage 26, usually overlooked by critics, is important in various respects. First, Aristotle does not conceive of the perception of a good as an *input,* a physical movement of a perceptual kind, which enters the mind of the agent from outside, causing a necessary chain of reactions. The perception of the end is an activity and not a movement. Second, the conclusion that "all actions would come to be forced" (b10-11) is evidently given as absurd by Aristotle: it is clear that not all actions are forced, and the social practice of rewards, punishments, and forgiveness suffices to attest this (1109b30-2). Aristotle seems to refer to the opinion of some anonymous interlocutor of deterministic tendencies, who might advance such an objection to his theory. This confirms that the argumentation of the *Nicomachean Ethics* takes place against the background of the discussions of the Platonic Academy.

Aristotle's analysis from chapter III 2 onwards focuses on the particular case of deliberate action and not on the voluntary in general. How so? Normally in his explanations Aristotle does not tend to give Fregean-type definitions, which determine exactly the scope of a set, and serve to establish clearly whether an entity *x* falls within a certain set *y* or does not: his are paradigmatic definitions, which illustrate the perfect case of an entity or event. This is a practice of Platonic origin, which has the advantage of clearly illustrating the defined object, but which leaves the boundaries of the defined set undetermined. Consistent with this approach in *NE* III 3-4, Aristotle illustrates the psychological mechanism related to choices about "what depends on us" by referring to deliberation, since deliberation is clearly, for him, the strongest sign that our actions depend on us; in fact, he says that "we deliberate about things that depend on us and are realizable" (1112a30-1).

Here lie the greatest difficulties for Aristotle himself. They relate to the relation between character and action, a theme we have seen discussed already in

the Academy (text 11), and which, as we already saw, calls into question the thesis that our choices depend on us.

We have seen that (1) for Plato both character and action are akousia; that, (2) for one of his opponents, character is akousion but action is hekousia. Aristotle also takes a stand on this problem: according to him, character is formed through the repetition of action, and

> 27. Is by choosing what is good or bad that we are men of a certain character, and not by simply holding some opinion (1112a1-2).

But since actions are voluntary, the character of each person, which is derived from the actions performed, is also a voluntary thing, even if it becomes at some point something necessary:

> 28. Actions and states of character (hexeis) are not voluntary (hekousioi) in the same way; for we are masters (kurioi) of our actions from the beginning until the end, if we know the particular facts, but though we control the beginning of our states the gradual progress is not obvious, any more than it is in illnesses. All the same, because it was in our power (eph'hēmin) to act in this way or not in this way, therefore the states are voluntary (hekousioi) (1114b30-1115a3).

In a sense, then, Aristotle stands somewhere between Plato and his opponent. However, some critics are convinced that, depending on character, all actions as well are necessary, although passage 28 seems to say the opposite, in fact it says "we are masters of our actions from the beginning until the end". Let us see how this is possible.

According to Aristotle, character determines the perception of the end, which serves as the final cause:

> 29. Absolutely and in truth the good is the object of wish, but for each person the apparent good; that that which is in truth an object of wish is an object of wish to the good man, while any chance thing may be so to the bad man, as in the case of bodies also the things that are in truth wholesome are wholesome for bodies which are in good condition, while for those that are diseased other things are wholesome—or bitter or sweet or hot or heavy, and so on; since the good man judges each class of things rightly, and in each the truth appears to him (1113a23-31).

The issue is further debated in NE III 5. In this passage Aristotle sets up a kind of dialogue, with questions, answers, and objections. It is divided into three parts of unequal extent: a brief demonstration in the positive that virtue and vice depend on us (1113b 8-13); a discussion of some possible objections to this thesis, a discussion that takes up most of the chapter (1113b1-1114b 21); and a general conclusion (1114b 21-1115a 3), which I have already quoted, passage 28.

Let us see the first two parts in succession. The argument has a very simple form, and is based on what has been established in the preceding chapters; Aristotle argues that:

> 30.
> – Good deeds and bad deeds are up to us.
> – Being good and bad consists (touto [...] einai) in doing the good deeds and the bad ones, therefore

– Whether we are good or bad is up to us (1113b 8-13).

In this demonstration there appears, as an explication of the first premise, a much discussed passage, in fact some consider it the passage in which Aristotle most clearly expounds an indeterministic view of human action, and others deny this:

> 31. Where it depends on us (eph'hēmin) to act it depends on us (eph'hēmin) also not to act, and in what [depends on us] the "no", the "yes" also [does it]; so that, if to act, where this is noble, depends on us (eph'hēmin), not to act, which will be base, will also depend on us (eph'hēmin), and if not to act, where this is noble, depends on us (eph'hēmin), to act, which will be base, will also depend on us (eph'hēmin) (1113b7-11).

I don't have time to explore all the problems of this disputed passage (see for a wider discussion, Natali 2023), but here Aristotle seems to me to confirm what he said in text 20.

In the discussion that follows Aristotle examines various objections, and the most important is the third, which takes up the argument of text 29, according to which we all follow what appears good to us. Aristotle repeats the thesis that the way in which the end is presented derives from one's character, and he refers back to the previous answer, for if one is the cause and responsible (aitios) for one's character one is also responsible for the way in which one perceives something as apparent good (1114a 31-b3).

On this answer the objections of the critics have concentrated. Aristotle seems to argue that it is true that character, once acquired, makes necessary, in a sense, the responses to be made to the stimuli of experience. Although character depends on actions, once acquired it no longer depends on us:

> 32. It is irrational to suppose that a man who does unjust acts does not wish to be unjust, or a man who does self-indulgent acts to be self-indulgent. But if without being ignorant a man does the things which will make him unjust, he will be unjust voluntarily (ekōn). Yet he cannot, if he wishes, to cease to be unjust and become just. For neither does the man who is ill can become well—although he had had the possibility to become ill voluntarily, through living incontinently and disobeying his doctors. In that case it was then open to him not to be ill, but not now, when he has neglected his health, just as when someone has let a stone go it is too late to recover it; but yet it depended on him (ep'autōi) to throw it, since the principle (archē) was in him. So, too, to the unjust and to the self-indulgent man it was open at the beginning not to become men of this kind, and so they are such voluntarily (hekontes); but now that they have become so it is not possible for them not to be so (1114a 11-21).

Modern critics have found in these passages arguments to object against an indeterministic interpretation of Aristotle's position. The critics argue, in a vaguely Platonic way, that since Aristotle also places a great deal of importance on education and the formation of character, one might think that only in youth, before the formation of a stable character, is man really capable of acting in opposite ways, but when character is formed this is no longer possible for him. A man with a particular character, e.g. a just man, they say, can only make

corresponding actions, the just ones. Actions depend on character and character depends on education, so basically actions would not depend on us.

To this several objections may be made. First, that it is doubtful whether education is really an efficient cause external to the agent. Because of that, the principle of passage 25 is in any case observed. Moreover, even if we admit that education is an efficient cause, it does not constitute a completely external cause, because we are a contributing cause to the development of our character, as Aristotle observes:

> 33. for we are ourselves somehow co-responsible (sunaitioi pōs) of our states of character (1114b23).

But the the criterion established in the passage 25 requires that we do not contribute *at all* to the character formation, if this has to be considered involuntary, so our character depends on us.

Again, it is not sure that character necessarily determines our actions. Sometimes Aristotle says that a just person can only act justly (*NE* 1129a14-17), but sometimes he admits the opposite possibility, possibly following an Academic theory (*Topics* IV*,* 126a36-b1: even god and the sage can do evil), and says that that one can perform unjust actions without being unjust (1137a20-26). On closer inspection, the passage 32 says the same. When saying: "If […] a man does the things which will make him unjust, he will be unjust voluntarily", Aristotle pictures a just man who does unjust acts, and *successively* becomes unjust. Since Aristotle admits that we can change our character over time, it is not a necessary determinant of our actions. In the *Categories* he describes with particulars the process by which a character change can happen:

> 34. For it is possible for the healthy to fall sick and for the white to become black and the hot cold; and it is possible to become bad instead of good or good instead of bad. For the bad man, if led into better ways of living and talking, would progress, if only a little, towards being better. And if he once made even a little progress it is clear that he might either change completely or make really great progress. For however slight the progress he made to begin with, he becomes ever more easily changed towards virtue, so that he is likely to make still more progress; and when this keeps happening it brings him over completely into the contrary state, provided time permits (13a20-31).

One could argue that the virtuous or the vicious for the most part performs virtuous or vicious actions, with at most a conditional necessity, in the sense that as long as an individual has a given character, he acts consistently with it most of the time, but that does not preclude the character from changing (*Politics* 1332b3-8; *Rhetoric* 1359a30-b5).

This whole discussion should not be seen in the light of contemporary discussion, in which compatibilism is seen as the *standard* philosophical position. As De Caro notes:

> 35. Per decenni […] il compatibilismo ha solcato trionfalmente le acque della filosofia angloamericana, rappresentando senz'altro il punto di vista maggioritario, se non egemone, rispetto alla questione della libertà (De Caro 2004: 69).

So one can see in today's critics a strong tendency to treat Aristotle as a proto-compatibilist.

In reality Aristotle's discussion starts from the Academic theses, which we have seen in passages 11 and 12. The Academic discussion seems to have focused on the relation between character (ethos) and action (praxis), on the relation between voluntariness (hekōn) and responsibility (aitios), with a great variety of solutions, and Aristotle's stance must be seen against that background. He wants to make both character and actions depend on us.

Another objection to the Aristotelian position, to which Aristotle replies in chapter III 5, takes up the Socratic idea that doing evil is the fruit of a miscalculation, and connects it with the question of the character. Aristotle seems to say that this position brings to admit an aristocratic conception of human nature:

> 36. If on the contrary no one is responsible for his own evildoing, but everyone does evil acts through ignorance of the end, thinking that by these he will get what is best—then the aiming at the end is not the result of a choice, but one must be born with an eye, as it were, by which to judge rightly and choose what is truly good, and he will be well endowed by nature who is well endowed with this (1114b3-8).

Aristotle's answers are important, because in him the question of responsibility is transformed: he no longer asks only whether the wicked are to blame for what they do, as in archaic poetry, but says that the same criterion must apply to all actions, both good and wicked. From the problem of guilt we thus pass to the problem of responsibility in general:

> 37. If this is true, then, how will virtue be more voluntary than vice? To both men alike, the good and the bad, things are the same: the end appears and is fixed by nature or however it may be, and it is by referring everything else to this that men do whatever they do. Whether, then, it is not by nature that the end appears to each man such as it does appear, but something also depends on him, or the end is fixed by nature but virtue is still voluntary because the good man does the rest voluntarily [if all this is true] vice also will be none the less voluntary as virtue (1114b12-20).

This thesis is consistent with the position expressed in text 26: the final cause does not necessitate the action of the moving cause, because if it were so all actions would be forced. In other words, whatever the origin of the conception of the end, the actions we perform in order to reach it are ours, because the first efficient cause of the movements of our body is us.

7. Alexander of Aphrodisias

Aristotle's school continued to discuss these matters for a long time. The most complete example are the works of Titus Aurelius Alexander, called Alexander of Aphrodisias, who lived between the second and third centuries A.D., i.e., more than five hundred years after Aristotle. He was the holder of a chair of Aristotelian philosophy in Athens, established by Marcus Aurelius at the expense of the imperial treasury, along with other chairs of Platonic, Stoic, and Epicurean philosophy. From his works we see that Alexander intended his task as a defence and actualization of Aristotle's doctrines: defence of Aristotle's

superiority over other philosophical schools, and actualization in the sense that in Aristotle one can find the tools to solve the philosophical problems of the present time, even if the master did not formulate them explicitly. His strategy resembles that of the Marxists or Thomists of the twentieth century, intent on actualizing the master's doctrines from within an institution dedicated to that end, and convinced that the best strategy consists in a return to the origins, to the text of the school's founder.

For this reason Alexander, in a writing *On destiny,* addressed to emperors Caracalla and Geta, proposes an Aristotelian theory of destiny (heimarmenē), even if in Aristotle the word is almost completely absent. It is true that it was a traditional theme, and was part of the philosophical debate from IV century B.C. onwards (cf. § 3). But from a philosophical point of view, the most important debate on destiny developed in Hellenistic age, with the Stoic and Epicurean schools. Alexander, in reference to those Hellenistic debates, set out to demonstrate that destiny exists, but it does not necessarily determine human actions, and that it must be understood on the basis of the Aristotelian theory of causes. Fate corresponds to the individual nature of each individual, in particular to his character, but character does not necessarily make us to act in a particular way. It influences our actions only for the most part, and there is always the possibility for the individual to react to his character, i.e. to act against his tendencies, on the basis of a rational deliberation.

It is possible that in his treatise Alexander repeats many of the arguments of a lost commentary by him on Aristotle's *De interpretatione*, Chpt. 9, a chapter which has a very polemical attitude towards determinism (cf. § 4). Alexander takes up such an attitude, and polemicizes against some determinists, who are never named but in whom it is easy to recognize the Stoic philosophers. The core of his anti-deterministic polemic is that man is the first cause of his own actions, and that there is nothing absurd about a frankly indeterministic perspective. Hence he goes so far as to assert that in two identical situations we always have the possibility of choosing one action or its opposite. It is sometimes said that in Alexander choice is not causally determined, but this does not appear to be true. What Alexander is saying is that choice is caused by us, because we are the first cause of our choices and actions:

> 38. The things that come to be in accordance with a cause do not always and in every case have the cause of their coming to be outside themselves. For it is on account of a freedom of this sort that something depends on us, because we ourselves are in control of the things that come about in this way and not some cause from outside. And for this reason the things that come about in this way do not come about without a cause having the cause from us. For man is the beginning and cause of the actions that come about through him and this is what being it is for a man, to have the principle of his acting in himself (*De fato* 15, 185, 11-17, transl. Sharples modified).

In fact, destiny corresponds to the individual's character and nature, and does not necessarily influence choices:

> 39. The things that come to be in accordance with nature do not do so of necessity but the coming to be of the things that come to be in this way can be sometimes hindered; for which reason the things that come to be in accordance with nature come to be for the most part but not of necessity […] it is from this, for the

> most part, that the lives of men and their deaths derive their patterns. At any rate we see that the body—through being like this or like in nature—is affected both in disease and in death in accordance with its natural constitution but not by necessity: for treatments and changes of climate and doctor's orders and advice from the gods are sufficient to break such a pattern. And in the same way in the case of the soul too one would find the choices and actions and ways of life of each individual differing from, and contrary to his natural constitution. For "men's character is their guardian spirit" according to Heraclitus, that is, their nature. For man's actions and life and endings can be seen to be for the most part in accordance with their natural constitutions and dispositions (*De fato* 11, 169, 28-170, 21, transl. Sharples modified).

8. Aristotle's Historical Position

Those who look at Aristotle from modern philosophical positions, such as contemporary ethical compatibilism, have frequently judged his position to be either confused or naive.

On the contrary, from the perspective of the ancient debate, it seems to me that Aristotle was a very sensible interlocutor in the philosophical discussion of his time. That discussion, which originated in the Academy, was an important moment in the ancient world's centuries-long discussion about guilt and responsibility, a conversation that began in the eighth century BC and lasted at least until the sixth century A.D., a span of more than 1,400 years.

In Aristotle, as already in Socrates, the debate on the responsibility and independence of the agent is detached from the sphere of the relationship with the gods and becomes an ontological problem, to be studied in moral psychology and in physics. This modification will characterize all ancient discussions on destiny for centuries, until the rise of Christianity, when the questions of divine grace and human salvation will again place the problem of determinism in a theological perspective, with new coordinates but, often, using ancient formulations.

From Aristotle onwards, the question of responsibility and causation is no longer merely that of the culpability of the wicked, but becomes the problem of human responsibility in general, for every action whether good or bad. The notion of "what depends on us", introduced by Aristotle became the central theme of the debate in the Hellenistic and imperial ages. The philosophers focused on how to understand this phrase, and on establishing under what conditions an act could be said to depend on us.

In that long debate in antiquity, questions were asked, distinctions formulated and concepts elaborated that have remained at the basis of European debate from the Middle Ages to the contemporary age. A survey of the ancient debate on responsibility can therefore be useful, both to show the origins and hidden implications of the terms of the modern debate, and to indicate possible alternatives to the unconscious assumptions of contemporary thought.

From the contemporary point of view, Aristotle seems to me particularly interesting because he proposes an original relationship between causality and determinism. In fact, for centuries the problem of determinism was presented in the following form: either to accept that everything is necessary, or to admit that there is a movement without a cause, that is, that there can be something that suddenly appears from nothing. Aristotle on the contrary formulated a very complex notion of causation, which does not always imply necessity. So every-

thing has a cause, but not everything is necessary, and archē taken in a strong sense is the key term of the Peripatetic position.

That such a move should be repeated even today was advocated a few years ago by P. Ricouer, according to whom only by revising our current notion of cause can we find a way of thinking about human free will without falling into unresolvable puzzles. In fact sometime Aristotle's critics seem to use an antiquated conception of causality to criticize him. This happens at a time in which the epistemological debate tends to revalue conceptions of cause as dependence, and not only as production, in a way similar to Aristotle's ones. There is, on my opinion, the possibility to study the Aristotelian theory of causality and of "what depends on us" not only as a historical curiosity, but as a source of possible new contemporary perspectives.[3]

References

Aristotle 1984, *The Complete Works of Aristotle*, *The Revised Oxford Translation*, Barnes, J. (ed.), Princeton: Princeton University Press.

Beebe, H., Hitchcock, C., & Menzies, P. (eds.) 2009, *The Oxford Handbook of Causation,* Oxford: Oxford University Press.

Bobzien, S. 1998, *Determinism and Freedom in Stoic Philosophy,* Oxford: Clarendon Press.

[3] This article touches on many questions, on each of which entire libraries have been written, so I have preferred not to burden the text with long notes, whose information would inevitably appear lacking. I will limit myself here to giving some basic references, of general works on the subject indispensable for a first information on these questions.

For an overview of theories of fate and guilt in poetry and tragedy, see Mondolfo 1958 and Magris's 2008, a volume which has a theoretical perspective of its own. On the legal context see Merker 2016, Pepe 2012 and 2019. On Greek intellectualism one can see Monteils-Laeng 2014. The essays collected in De Luise and Zavattero 2019 discuss determinism from Homer to William of Ockam. The volume edited by De Caro, Mori and Spinelli 2014 gives a general information about the philosophical debate of the ancient and modern philosophical schools. In English on these issues see now Frede 2011, who replaced Dihle 1982. For a very different reconstruction from the one presented here see the works of Bobzien 1998, 1998a, 2014. On *De interpretatione* see Gaskin 1995 and Whitaker 1996. On the expression "what depends on us" a good general overview is in Eliasson 2008, see also Labarrière 2009, the volume edited by Destrée, Salles and Zingano 2014, and Natali 2023. On EE II 6, cf. Kenny 1979 and Natali 2004, Chpt. 9. On issues of character and choice see Furley 1967, Donini 1986, Di Muzio 2000 and Meyer 2011. A typical compatibilist reading of Aristotle is Everson 1990. A rather indeterminist reading instead has been proposed, with different arguments, by Sorabji 1980, Broadie 1991, Di Muzio 2008 and Destrée 2011. On the contingent and the accidental in *Metaphysics* and *Physics* see Rossi 2011 and Masi 2015. Historians of Stoicism generally tend to defend the Stoics against Alexander's criticism, see Long 1970 and Sharples 1975, as well as Bobzien cited above. On Alexander see now Harari 2023. On modern compatibilism see De Caro 2004; Ricoeur's view is in Ricoeur 1990; the early twenty-first century debate on causality is well presented in the volume edited by Beebe, Hitchcock & Menzies 2009. Aristotle's *Nicomachean Ethics* and Alexander's *De fato* have been translated into Italian and commented, among others, by Natali 1999 and 2009; Aristotle's *Eudemian Ethics* by Donini 1999.

Bobzien, S. 1998a, "The Inadvertent Conception and Late Birth of the Freewill Problem", *Phronesis*, 43, 133-75.

Bobzien, S. 2014, "Choice and Moral Responsibility (NE iii 1-5)", in Polansky, R. (ed.), *The Cambridge Companion to Aristotle's Nicomachean Ethics*, Cambridge: Cambridge University Press, 81-109.

Broadie, S. 1991, *Ethics with Aristotle,* Oxford: Oxford University Press.

De Caro, M. 2004, *Il libero arbitrio: Una introduzione*, Roma-Bari: Laterza.

De Caro, M. Mori, M., & Spinelli, E. (eds.) 2014, *Libero arbitrio: Storia di una controversia filosofica,* Roma: Carocci.

De Luise, F. & Zavattero, I. (eds.) 2019, *La volontarietà dell'azione tra Antichità e Medioevo,* Trento: Università degli Studi di Trento.

Destrée, P. 2011, "Aristotle on Responsibility on One's Character*",* in Pakaluk, M. & Pearson, G. (eds.), *Moral Psychology and Human Action in Aristotle,* Oxford: Oxford University Press, 285-318.

Destrée, P., Salles, R., & Zingano, M. (eds.) 2014, *What is Up to Us? Studies on Agency and Responsibility in Ancient Philosophy*, Sankt Augustin: Academia Verlag.

Dihle, A. 1982, *The Theory of Will in Classical Antiquity,* Berkeley-London: California University Press.

Di Muzio, G. 2000, "Aristotle on Improving One's Character", *Phronesis*, 45, 205-19.

Di Muzio, G. 2008, "Aristotle's Alleged Determinism in the Nicomachean Ethics", *Journal of Philosophical Research,* 33, 19-32.

Donini, P.L. 1986, *Ethos: Aristotele e il determinismo*, Alessandria: Edizioni dell'Orso, English translation: Louvain-la-Neuve: Peeters, 2010.

Donini, P.L. (ed.) 1999, *Aristotele: Etica Eudemia*, Roma-Bari: Laterza.

Eliasson, E. 2008, *The Notion of* That Which Depends on Us *in Plotinus and Its Background,* Leiden-Boston: Brill.

Everson, S. 1990, "Aristotle's Compatibilism in the *Nicomachean Ethics*", *Ancient Philosophy*, 10, 81-103.

Frede, M. 2011, *A Free Will: Origin of the Notion in Ancient Thought,* Long, A.A. (ed.), Berkeley: University of California Press.

Furley, D.J. 1967, "Aristotle on the Voluntary", in Id., *Two Studies on Greek Atomists*, Princeton: Princeton University Press, 171-238.

Gaskin, R. 1995, *The Sea Battle and the Master Argument: Aristotle and Diodorus Cronus on the Metaphysics of Future,* Berlin-New York: De Gruyter.

Harari, O. 2023, "Alexander of Aphrodisias' Theory of Action and the Capacity of Doing Otherwise", *Apeiron*, DOI: https://doi.org/10.1515/apeiron-2022-0114.

Homer 1919, *Odyssey*, Murray, A.T. (transl.), Loeb Classical Library, Cambridge (MA)-London: Harvard University Press.

Homer 1924, *Iliad*, Murray, A.T. (transl.), Loeb Classical Library, Cambridge (MA)-London: Harvard University Press.

Kenny, A. 1979, *Aristotle's Theory of the Will,* London: Duckworth.

Labarrière, K.L. 2009, "De 'ce qui dépend de nous'", *Les Études philosophiques,* 88, 5-24.

Long, A.A. 1970, "Stoic Determinism and Alexander of Aphrodisias *De fato* (I-XIV)", *Archiv für Geschichte der Philosophie*, 52, 247-68.

Magris, A. 2008, *Destino, provvidenza, predestinazione: Dal mondo antico al cristianesimo,* Brescia: Morcelliana.

Masi, F.G. 2015, *Qualsiasi cosa capiti: Natura e causa dell'ante accidentale. Aristotele* Metafisica *Epsilon 2-3*, Amsterdam: Hakkert.

Merker, A. 2016, *Le principe de l'action humaine selon Démosthène et Aristote,* Paris: Les Belles Lettres.

Meyer, S.S. 2011, *Aristotle on Moral Responsibility,* Oxford: Oxford University Press, 2nd edition.

Mondolfo, R. 1958, *La comprensione del soggetto umano nell'antichità classica,* Firenze: La Nuova Italia.

Monteils-Laeng, L. 2014, *Agir sans vouloir: Le problème de l'intellectualisme moral dans la philosophie ancienne,* Paris: Classiques Garnier.

Natali, C. (ed.) 1999, *Aristotele: Etica Nicomachea*, Roma-Bari: Laterza.

Natali, C. 2004, *L'action efficace : Études sur la philosophie de l'action d'Aristote,* Louvain-la-Neuve: Peeters.

Natali C. (ed.) 2009, *Alessandro d'Afrodisia: Il destino. Trattato sul destino e su ciò che dipende da noi, dedicato agli imperatori*, Sankt Augustin: Academia Verlag, 2nd edition.

Natali, C. 2023, "L'expression 'ce qui dépend de nous' chez Aristote : origine et importance", *Revue de philosophie ancienne,* 41, 7-54.

Pepe, L. 2012, Phonos*: L'omicidio da Draconte all'età degli oratori,* Milano: Giuffré.

Pepe, L. 2019, *Atene a processo: Il diritto ateniese attraverso le azioni giudiziarie,* Bologna: Zanichelli.

Plato 1997, *Complete Works*, Cooper, J.M. & Hutchinson, D.S. (eds.), Indianapolis-Cambridge: Hackett.

Plutarch 1916, *Alcibiades and Coriolanus*, Perrin, B. (transl.), Loeb Classical Library, Cambridge (MA)-London: Harvard University Press.

Ricoeur, P. 1990, *Soi-même comme un autre,* Paris : Seuil.

Rossi, G. 2011, *El azar según Aristóteles: Estructuras de la causalidad accidental en los procesos naturales y en la acción,* Sankt Augustin: Academia Verlag.

Sharples, R.W. 1975, "Aristotelian and Stoic Conceptions of Necessity in the *De fato* of Alexander of Aphrodisias", *Phronesis*, 20, 247-74.

Sharples, R.W. 1983, *Alexander of Aphrodisias: On Fate*, Sharples, R.W. (transl.), London: Duckworth.

Sorabji, R. 1980, *Necessity, Cause and Blame: Perspectives on Aristotle's Theory*, London: Duckworth.

Whitaker, C.W.A. 1996, *Aristotle's* De interpretatione*: Contradiction and Dialectic*, Oxford: Clarendon Press.

Spinoza on Freedom, Feeling Free, and Acting for the Good

Leonardo Moauro

University of California, San Diego

Abstract

In the *Ethics*, Spinoza famously rejects freedom of the will. He also offers an *error theory* for why many believe, falsely, that the will is free. Standard accounts of his arguments for these claims focus on their efficacy against *incompatibilist* views of free will. For Spinoza, the will cannot be free since it is determined by an infinite chain of external causes. And the pervasive belief in free will arises from a structural limitation of our self-knowledge: because we are aware of our actions but unaware of their causes, we suppose that we alone must be responsible for them. Yet I argue that the standard accounts miss a further element of Spinoza's arguments that also targets *compatibilist* views on which free will is consistent with a specific kind of determination—namely, self-determination in accordance with our *value judgments*. For Spinoza, we are misled not only in supposing that our actions lack external determination but also in thinking that they are determined by our representations of value. In fact, our actions are determined by our appetites, which are blind to our value judgments. And the pervasive belief that our actions are determined by such judgments arises from the *projection* of value onto the objects we seek. As he denies us free will, then, Spinoza also denies us a capacity central to agency—the capacity to determine our actions in accordance with our ideas of the good. This makes his arguments against free will more consequential, and more radical, than commonly assumed.

Keywords: Spinoza, Free will, Projectivism, Error theory, Agency.

1. Introduction

At first glance, Spinoza's rejection of free will appears to be among the most straightforward doctrines of the *Ethics*. Spinoza states flatly that "[i]n the Mind there is no absolute, *or* free, will, but the Mind is determined to will this or that by a cause which is also determined by another, and this again by another, and so on to infinity" (2p48 | G II/129: 483).[1] The will is not free for Spinoza because

[1] All translations of Spinoza's works are drawn from Spinoza 2016 and include citations of Spinoza 1925. Unless otherwise noted, all references are to Spinoza's *Ethics*, for which

Argumenta 11, 1 (2025): 35–50
ISSN 2465-2334

DOI 10.14275/2465-2334/20230.moa
First published: 19 June 2023

acts of willing are in every case determined by an infinite chain of causes. The question of human freedom is thus quickly settled—we have no free will.

Yet a second look at this rejection reveals two further issues of considerable interest. First, Spinoza does not simply deny us free will; he expresses interest in why many *believe*, falsely, that their wills are free. He attributes the pervasiveness of this false belief to a deeply rooted but illusory *experience* of our actions as free. I will call this the *feeling of freedom*.[2] Because many take this feeling as evidence that their actions are freely willed, Spinoza provides an *error theory* to strengthen his case against free will: "[M]en are deceived in that they think themselves free [NS: i.e., they think that, of their own free will, they can either do a thing or forbear doing it], an opinion which consists only in this, that they are conscious of their actions and ignorant of the causes by which they are determined. This, then, is their idea of freedom—that they do not know any cause of their actions" (2p35s | G II/117: 473). The feeling of freedom arises from a structural limitation of our self-knowledge. We are aware of our actions but unaware of the causes that determine them. Since we lack this knowledge, it seems to us that our actions follow from our wills alone, and thus that our wills are free.[3]

The second issue concerns the *kind* of free will Spinoza denies us. 2p48 most clearly targets *incompatibilist* conceptions of freedom that analyze it as an absolute faculty of choice—that is, an undetermined, two-way power to do or not do something. For incompatibilists, the will is free just in case, holding all conditions fixed, it can order or forbear from ordering an action. Incompatibilist accounts of free will were defended most prominently by late scholastics like Suárez and Molina, and by Bishop Bramhall in a famous dispute with Hobbes. Descartes also defended incompatibilist freedom, albeit in a qualified form.[4]

But *compatibilist* accounts that make free will consistent with determination were also prominent in the early modern period. Consider Leibniz's view of freedom as stated in the *Theodicy*, on which the will is free because it is *self*-determining in accordance with our representations of value:

I use the following conventional abbreviations: First number=Part, 'pref'=preface, 'app'=appendix, 'D'=definition, 'p'=proposition, 'd'=demonstration, 'c'=corollary, and 's'=scholium. So, for example, '2p48' refers to '*Ethics* Part 2, proposition 48'. For Spinoza's other works, I use the following standard abbreviations: 'KV'=*Short Treatise on God, Man, and his Well-Being*, 'Ep'=Letter.

[2] I take this useful phrase from Barry 2016. It is critical to underscore the difference between the *feeling* of freedom and the *belief* that we have free will. As I use the terms, a 'belief' indicates a higher level of cognition than a 'feeling', and what we feel is not always responsive to what we believe. For instance, we can imagine that Spinoza himself felt that his actions were free in the sense at issue here, even if he rejected free will. See Barry 2016: 632-33 for more.

[3] See also 3p2s, 4pref, 5p5d. For accounts of this argument, see Barry 2016 and Melamed 2017. Spinoza's interest in the feeling of freedom is also noticed by Bennett (1984: 170), Deleuze (1988: 20), Klein (2003: 154), Nadler (2006: 186), and Sharp (2007: 740).

[4] For Molina's views on free action, see *Concordia* IV and the introduction to Molina 1988. For Suárez's views, see *Disputationes Metaphysicae* XVII-XXII and Penner 2013. For Descartes' view of free will, see the Fourth Meditation (CSM II 40), the Second Set of Replies (CSM II 116), and the *Passions of the Soul*, I.41 (CSM I 343). For commentary that compares Descartes' view with Molina's and Suárez's views, as well as with Leibniz's, see Wee 2006. Finally, for a very helpful overview of the various positions on free will defended by key figures in seventeenth century Europe, see Ragland 2017.

> [Freedom] consists in *intelligence*, which involves a clear knowledge of the object of deliberation, in *spontaneity*, whereby we determine, and in *contingency*, that is, in the exclusion of logical or metaphysical necessity. Intelligence is, as it were, the soul of freedom, and the rest is as its body and foundation. The free substance is self-determining and that according to the motive of good perceived by the understanding, which inclines it without compelling it: and all the conditions of freedom are comprised in these few words (*Theodicy* §288: 306, emphases mine).

For Leibniz, an action is free just in case it is *spontaneous*, *contingent*, and *intelligent*—that is, just in case it is self-determined according to a representation of value, and it is not metaphysically necessary. On this account, freedom is not only consistent with but *requires* a particular kind of determination. Indeed, Leibniz calls intelligence the "soul of freedom". For him, free will involves a paradigmatic exercise of practical reason: the selection of an object or course of action based on some normative demand, like the object's goodness or rightness. Our capacity to order our actions in accordance with such demands is makes our will free on compatibilist views, even if determined.

These issues raise two related questions about Spinoza's engagement with the problem of free will. First, does Spinoza's rejection of free will only target incompatibilist views like Descartes', or does it also target compatibilist views like Leibniz's? Second, if the rejection does extend to compatibilist views, does Spinoza's error theory—his account of the *feeling* of freedom—also extend to the belief that our actions are determined by representations of value, which compatibilist views identify as the key condition of freedom?

In this paper, I argue that both Spinoza's rejection of free will and his account of the feeling of freedom *do* extend to compatibilist views like Leibniz's.[5] In the first instance, this is because Spinoza denies that we satisfy Leibniz's contingency condition. Spinoza agrees with Leibniz that freedom is consistent with self-determination. A thing is free just in case it exists and acts by its nature alone (1D7).[6] Yet for Spinoza, this implies that everything exists and acts according to strict metaphysical necessity. Indeed, Galen Barry has persuasively argued that Spinoza's error theory includes an explanation of the *feeling of contingency*—a deeply rooted but misleading experience that our actions are not metaphysically necessary, that we could have done otherwise.

But I will argue that Spinoza's rejection of free will and his account of the feeling of freedom both go further than this. For Spinoza also denies that we satisfy Leibniz's intelligence condition, and his error theory includes an account of

[5] To be clear, I do not argue that Spinoza's arguments *succeed* against compatibilist views like Leibniz's. I only aim to show that Spinoza would also have rejected conceptions of free will that locate freedom not in the absence of determination but in the capacity to be determined by representations of value.

[6] It follows for Spinoza that only God is free, for only God acts by the necessity of his nature alone (1p17c2). In other words, Spinoza denies that we satisfy Leibniz's spontaneity condition. This point highlights that there are really *two levels* of contrast between Spinoza and Leibniz here. The first level is *conceptual*: What does free will consist in? The second level is *extensional*: Do human beings have free will? Spinoza and Leibniz both hold that freedom requires spontaneity. Leibniz adds to this that freedom also requires contingency and intelligence, and that human beings have free will in this full sense. Spinoza insists that while these conditions capture the ordinary conception of free will, it is neither the case that we enjoy this kind of freedom nor that this is what freedom, properly understood, consists in.

the *feeling of intelligence*—a deeply rooted but false experience of our actions as determined by representations of value. This is significant for several reasons. In the first place, Spinoza's denial of the intelligence condition and the account of the feeling of intelligence complete his rejection of free will. But these claims also have major implications for his conception of human agency. By denying us the "soul of freedom," Spinoza strikingly eliminates a key feature of many standard accounts of agency: the capacity to determine our actions in accordance with our view of the good. For Spinoza, our self-conception as beings sensitive to practical reasons is a defining but ultimately mistaken feature of human experience.

2. Freedom and the *Feeling* of Freedom

Freedom is one of the first notions Spinoza defines in the *Ethics*. For Spinoza, a thing is free just in case it is entirely self-determined: "That thing is called free which exists from the necessity of its nature alone, and is determined to act by itself alone. But a thing is called necessary, or rather compelled, which is determined by another to exist and to produce an effect in a certain and determinate manner" (1D7 | G II/46: 409). Immediately clear in this definition is Spinoza's denial of an incompatibilist conception of freedom as an undetermined power of choice. For Spinoza, freedom is opposed not to mere determination but to *compulsion*—that is, to determination by external causes. To be free is to be determined by one's nature alone.

Working with this conception of freedom, Spinoza swiftly puts to rest the question of human free will. Freedom can be ascribed only to God, "[f]or God alone exists only from the necessity of his nature (by 1p11 and 1p14c1), and acts from the necessity of his nature" (1p17c2 | G II/61: 425).[7] Only God, the one infinite and eternal substance, is self-determining. Human beings are part of an infinite chain of causes that determine one another to exist and produce effects: "Every singular thing, *or* any thing which is finite and has a determinate existence, can neither exist nor be determined to produce an effect unless it is determined to exist and produce an effect by another cause, which is also finite and has a determinate existence" (1p28 | G II/69: 432). Since the mind is determinate—since its existence and actions are both determined by things external to it—it is not free: "The Mind is a certain and determinate mode of thinking (by 2p11), and so (by 1p17c2) cannot be a free cause of its own actions" (2p48d | G II/129: 483).

For Spinoza, then, there is no doubt that our wills are not free. Even so, he frequently remarks on the pervasiveness of the *belief* in freedom, a belief arising from what he considers a deep-seated but misleading experience of our actions as freely willed. No matter what abstract argument against free will we accept, Spinoza thinks, we cannot help but feel that our actions are free. This *feeling of freedom* is a distinctive phenomenon that Spinoza takes it upon himself to explain, and for good reason. Since many take it as evidence for the claim that we *are* free,

[7] In the last two Parts of the *Ethics*, however, it seems that Spinoza *does* allow that human beings can achieve a kind of freedom that represents an ethical ideal. This freedom has to do with what Spinoza calls 'Strength of Character' (3p59s, 4p73s), which denotes actions that we undertake insofar as our mind understands or 'acts' in a technical sense (3p1, 3p3)—that is, insofar as it is self-determined (3D1, 3D2). This kind of freedom is exemplified by the figure of the 'free man' (4p67-73) and is the subject of *Ethics* 5, titled 'on Human Freedom'. For recent commentary on the relation between this ethical sense of freedom and the strict sense defined in *Ethics* 1, see Kisner 2019.

Spinoza's argument against free will will be all the more convincing if he can explain the feeling away. If we have no free will, then why does it seem to us that we do?

Spinoza's clearest answer to this question posits a structural limitation of our self-knowledge—our *ignorance* of the causal origin of our actions: "[M]en are deceived in that they think themselves free [NS: i.e., they think that, of their own free will, they can either do a thing or forbear doing it], an opinion which consists only in this, that they are conscious of their actions and ignorant of the causes by which they are determined" (2p35s | G II/117: 473). Since we are aware of our actions but have no idea of the external causes that determine them, it seems to us that they *lack* external causes, and are thus free. Using Leibniz's language, we might call this the *feeling of spontaneity*.[8]

Most scholars assume this is all Spinoza has to say about the feeling of freedom. But this is not so. For Spinoza recognizes that the *sense* of freedom that many people ascribe to themselves and to God includes more than lack of external determination. Indeed, he acknowledges that many will balk at his own account of divine freedom as complete self-determination "because they have been accustomed to attribute another freedom to God, far different from that we have taught (1D7), viz. an absolute will" (1p33s2 | G II/74: 437). An absolute will does not merely lack determination by external causes—it is also *contingent*: "By God's power ordinary people understand God's free will and his right over all things which are, things which on that account are commonly considered to be contingent" (2p3s | G II/87: 449). For ordinary people, in other words, God's freedom implies that his actions are not metaphysically necessary.

Of course, Spinoza believes ordinary people are mistaken. God could not have acted otherwise. While God alone is free in the sense of self-determination, God does not have *free will*: "God does not produce any effect from freedom of the will" (1p32c1 | G II/73: 435).[9] But the error of ordinary people is nonetheless indicative of something important about the feeling of freedom. When people think themselves and God as free, they have in mind the idea that they could have done otherwise. For instance, many feel that they have a capacity to suspend judgment: "experience seems to teach nothing more clearly than that we can suspend our judgment so as not to assent to things we perceive […] that the will, *or* faculty of assenting, is free" (2p49s | G II/132-33: 487).[10] From their own

[8] With the caveat that Spinoza does not specify here whether the feeling at issue is that our actions are *undetermined* or that they are *self*-determined. Spontaneity as Leibniz thinks of it consists in self-determination.

[9] This explains why it is not contradictory for Spinoza to say both (i) that God alone is free and (ii) that God has no free will. These two claims employ two different conceptions of freedom. Indeed, Spinoza seems to use the phrase 'free will' to indicate a richer conception of freedom that includes contingency in addition to self-determination (and, as I will later argue, intelligence). Thus, God is free because he is self-determined, but he has no free will because his actions are not contingent.

[10] Spinoza likely has Descartes in mind here, who equates free will with the power to grant or withhold assent to a proposition or to a proposed course of action. See *Principles of Philosophy* I.39: "That there is freedom in our will, and that we have power in many cases to give or withhold our assent at will, is so evident that it must be numbered among the first and most common notions that are innate in us" (CSM I: 205-206). The cases in which we do *not* have absolute power to give or withhold assent, Descartes makes clear in a May

experience, many hold that it is possible to withhold assent from a particular idea or course of action, which implies that their actions are not metaphysically necessary. In other words, they believe their actions are free because of their *feeling of contingency*.

When Spinoza rejects human free will, then, he is also denying that our actions are contingent. Note that this point complicates the standard reading of Spinoza's account of the feeling of freedom. Can the account of the feeling of freedom in 2p35s, taken on its own, *also* explain the feeling of contingency? Galen Barry (2016: 163-69) argues that it cannot. For Barry, Spinoza takes the feeling of contingency to arise from a distinctive kind of cognitive limitation tied to the mental phenomena of *association* and *vacillation*. Association is a psychological process by which the mind moves between two representational contents that are not logically connected. For Spinoza, this occurs when the body has been affected concurrently by different external stimuli—like the sound of a word and the sight of an object. To use Spinoza's example, the mind of a Roman citizen associates the auditory representation '*pomum*' with the visual representation of an apple (2p18s), moving seamlessly between the two.

Yet the mind often associates ideas with incompatible representational contents (2p44s). For instance, because I sometimes have breakfast at home and other times have breakfast out, my mind associates mornings both with cereal at home and with a cornetto and cappuccino *al bar*. Since the content of these ideas is logically incompatible, my mind cannot consider both at the same time. As a result, it switches rapidly or 'vacillates' between them, representing now the one and now the other. Crucially, since for Spinoza the mind is apt to confuse the properties of objects with the way those objects affect us (2p17s), the mind confuses its own vacillation between two representational contents with a property of the represented objects. That is, the mind *projects* its own vacillation between two objects onto the objects themselves, which it thus represents as contingent.[11] Finally, since our actions are often accompanied by representations of other, logically incompatible courses of action, we represent our own actions as contingent. We feel that we could have acted otherwise.

Barry offers a convincing explanation of how Spinoza accounts for the feeling of contingency. And of course, Spinoza denies contingency in nature. Since all things follow from the necessity of God's nature, all things occur with strict metaphysical necessity (1p16, 1p33). So, Spinoza denies that we satisfy Leibniz's second condition for freedom, and he explains why many nevertheless believe falsely that our actions are in fact contingent.

Yet Spinoza's rejection of free will goes further than this. For the ordinary conception of free will holds not only that we could have done otherwise but also that we act on the basis of *reasons*. This is why Leibniz's account of freedom centers the idea that our actions are determined by our representations of value—that

1644 letter to Mesland, are those in which the object of the will is clearly and distinctly perceived: "[I]f we see very clearly that a thing is good for us, it is very difficult—and on my view, impossible, as long as one continues in the same thought—to stop the course of our desire [to pursue]" (CSMK III: 233-34). According to Wee (2006: 395-98), Descartes holds that, in such cases, it is still possible for the agent to do otherwise by shifting her attention away from the clear and distinct perception.

[11] I say more about the phenomenon of projection in Section 3, where I argue that for Spinoza good and evil are also projected properties, and that this partly explains why we feel that our actions are intelligent.

they are *intelligent*. This, for Leibniz, is the "soul of freedom". A free action results from an exercise of practical reason, which involves selecting an option or a course of action due to some normative demand associated with it, like its goodness or rightness. But I argue that Spinoza denies that we satisfy this intelligence condition, and I also argue that he offers an error theoretic account of the *feeling of intelligence*—the deeply rooted but misleading sense that our actions are determined by representations of value.[12]

3. Spinoza on the Feeling of Intelligence

Spinoza's interest in the feeling of intelligence emerges in his account of what he calls 'decisions of the mind'—roughly, our choices to pursue one course of action over another. Spinoza introduces this notion at the beginning of *Ethics* 3, where he rejects Descartes' view that the mind determines the body to move via free acts of will. The feeling that our words and actions always follow from our decisions, Spinoza notes, is often cited as a point of support for the Cartesian view. Though they know not *how* the mind moves the body, many people claim that "they know by experience, that it is in the Mind's power alone both to speak and to be silent, and to do many other things which they therefore believe depend on the Mind's decision" (3p2s | G II/142: 495).

For Spinoza, the experience of our actions as dependent on our decisions is closely related to the feeling of intelligence. In an early work, the *Short Treatise on God, Man, and His Well-Being*, he identifies the mind's decisions with value judgments: "[B]efore our Desire extends externally to something, a decision has already taken place in us that such a thing is good. This affirmation, then, or taken generally, the power of affirming and denying, is called the Will" (KV II.16 | G I/80: 121). Spinoza claims here that a decision is an affirmation that some object is good or evil—that is, a value judgment.[13] Decisions lie at the foundation of all our actions. We judge something to be valuable and so desire and pursue it. To experience our actions as dependent on our decisions, then, is to experience them as determined by a value judgment.

Now, as we will see momentarily, the mature moral psychology Spinoza defends in the *Ethics* differs from this picture in fundamental ways. Most importantly, Spinoza argues in the later work that our actions are *not* in fact determined by value judgments—that they are not intelligent. Even so, Spinoza retains the idea that we *feel* that our actions are intelligent in that we experience them as depending on our representations of value. And as we will see in Sections 3 and 4, this feeling arises

[12] What Barry has to say about Spinoza's engagement with the intelligence claim is unsatisfying: "All mental activity, including decisions, is the result of reasons; and when we are conscious of those reasons, an action is felt as intelligent" (Barry 2016: 640). This brief explanation does not respect the difference between *normative* reasons—those consisting in a normative demand for action—and *explanatory* reasons—those that specify the causes of events. The feeling of intelligence is the feeling that our actions are determined by normative reasons. But it is unclear how this feeling might arise simply from consciousness of the reasons that explain our mental activity, since these are explanatory reasons.

[13] Earlier, I articulated the intelligence condition in terms of *representations* of value, not affirmations or judgments. But for Spinoza, these all amount to the same. In the *Ethics*, he identifies the will and the intellect (2p48-49). So, for Spinoza, to represent something is the same as to affirm something of it as true—that is, to make a judgment about it.

from the fact that the mind projects onto objects of desire the value properties which it then falsely takes to determine our actions.

We can begin by taking note of something beyond dispute. Spinoza claims in the *Ethics* that many believe that their actions depend on freely made decisions of the mind, and that this belief is *false*: "[T]he drunk believes it is from a free decision of the Mind that he speaks the things he later, when sober, wishes he had not said. So the madman, the chatterbox, the child, and a great many people of this kind believe they speak from a free decision of the Mind, when really they cannot contain their impulse to speak" (3p2s | G II/143: 496).[14] The reason this belief is mistaken is that the mind's decisions are in fact reducible to its *appetites*: "[E]xperience itself, no less clearly than reason, teaches that men believe themselves free because they are conscious of their own actions, and ignorant of the causes by which they are determined, *that the decisions of the Mind are nothing but the appetites themselves*, which therefore vary as the disposition of the Body varies. For each one governs everything from his affect" (3p2s | G II/143: 496-97, emphasis mine).

Spinoza thus claims that our mind's decisions are not free because they are nothing more than our appetites, which alone determine all our actions.[15] To understand why this analysis of decision might be plausible, we must consult the fundamental principle of Spinoza's moral psychology, the *conatus* doctrine: "Each thing, as far as it can by its own power, strives to persevere in its being" (3p6 | G II/146: 498), and this striving "is nothing but the actual essence of the thing" (3p7 | G II/146: 499). The actual essence of each thing consists in a power to persevere in existence, and our decisions and appetites must both be understood in terms of it: "When this striving is related only to the Mind, it is called Will; but when it is related to the Mind and Body together, it is called Appetite. This Appetite, therefore, is nothing but the very essence of man, from whose nature there necessarily follow those things that promote his preservation. And so man is determined to do those things" (3p9s | G II/147: 500).[16] So, for Spinoza, our decisions are identical to our appetites simply because both are different ways of considering the same, fundamental striving to persevere in being, which alone determines all our behavior.[17]

[14] See also the well-known letter to Schuller: "This is that famous human freedom everyone brags of having, which consists only in this: that men are conscious of their appetite and ignorant of the causes by which they are determined. So the infant believes that he freely wants the milk; the angry boy that he wants vengeance; and the timid, flight. Again, the drunk believes it is from a free decision of the mind that he says those things which afterward, when sober, he wishes he had not said. Similarly, the madman, the chatterbox, and a great many people of this kind believe that they act from a free decision of the mind, and not that they are set in motion by an impulse" (Ep. 58 | G IV/266: 428).

[15] 3p2s says that our *affects* determine our actions, but Spinoza later clarifies that appetites and affects are the same thing (3p9s). See Rutherford 2019 for an account of the relation between decisions of the mind and the affects.

[16] What Spinoza refers to as 'Will' here is identical to our decisions: "[T]he decision of the Mind and the appetite and the determination of the Body by nature exist together—or rather are one and the same thing" (3p2s | G II/144: 497).

[17] Of course, we are *also* always determined by external causes. But this external determination itself runs through our essential striving. Thus, for Spinoza, our actions are determined both by our essential striving alone and by the causal influence on this striving of external things (see e.g., 3D1-2 and 4p4).

How does this help us understand Spinoza's claim that we are mistaken in regarding our mind's decisions as free? On the one hand, it is clear that our decisions cannot be *freely made* in the sense that they cannot be spontaneous. For they are nothing but our affects which, as Spinoza claims, "vary as the disposition of the body varies". Our decisions, like the affections of our body, are always determined by external causes. We have no undetermined power of choice.

Yet later in *Ethics* 3, Spinoza points to a second, more fundamental way we are deceived about the decisions of the mind. As we saw, Spinoza associates decisions with judgments of value in the *Short Treatise*, claiming that these judgments determine our desires. Yet in the *Ethics*, he strikingly inverts this order of explanation. He claims that it is not our judgments of value that give rise to our desires but rather our desires that give rise to our judgments of value: "From all this, then, it is clear that we neither strive for, nor will, neither want, nor desire anything because we judge it to be good; on the contrary, we judge something to be good because we strive for it, will it, want it, and desire it" (3p9s | G II/148: 500).[18] For Spinoza, all our actions are determined by our appetites, which are expressions of our essential striving to persevere in being. This striving, he now tells us, is not determined by but instead determines our judgments of value. It follows that our actions are not determined by judgments of value. Holding something to be good is the *result* of our appetite for it, not an input that determines this appetite.

This point is critical to grasping Spinoza's rejection of the intelligence condition. An action is intelligent, recall, if it is determined by a representation of value. But for Spinoza, what we hold to be valuable is fixed by our strivings, appetites, or affects, and it is these alone that determine all we do.[19] By reducing the mind's decisions to appetites, then, Spinoza reveals something crucial about how these decisions are related to our actions. We may feel that our decisions lead us to act because they consist in value judgments. But Spinoza denies that our actions depend on our value judgments—our actions are determined entirely by our affects or appetites, which are expressions of an essential striving to persevere in being that is insensitive to representations of value. Thus, the feeling of intelligence is illusory. Our actions are never determined by our representations of value; instead, both are determined by our affects or appetites.

4. Spinoza's Value Projectivism

For Spinoza, I've argued, our actions are not determined by our value judgments but rather by our appetites, which alone govern all our behavior. But this point alone does not explain why we *feel* that our actions are determined by our value judgments—that is, that they are intelligent. Spinoza explains the feelings of spontaneity and contingency by appeal to necessary limits of our cognition. How might he explain the feeling of intelligence?

[18] For Spinoza, desire and appetite are the same thing (cf. 3p9s, 3DA).

[19] Spinoza later provides a helpful illustration of this claim "So each one, from his own affect, judges, *or* evaluates, what is good and what is bad, what is better and what is worse, and finally, what is best and what is worst. So the Greedy man judges an abundance of money best, and poverty worst. The Ambitious man desires nothing so much as Esteem and dreads nothing so much as Shame" (3p39s | G II/170: 516-17). The greedy man is one whose dominant appetites are for wealth (3DA XLVII). These appetites determine him to seek wealth *and* to judge that wealth is good. So, the value judgment is codetermined with his actions—it does not give rise to them.

I believe Spinoza offers an account of the feeling of intelligence on which the phenomenon of projection plays a key role.[20] As we've seen, Barry argues that, for Spinoza, we feel that our actions are contingent because the mind mistakes its vacillation between different ideas for a property that the objects represented by the ideas have in themselves—namely, their contingency. Similarly, I propose, Spinoza holds that we experience our actions as determined by value judgments because the mind mistakes its own appetites for properties things have in themselves—namely, being good or evil. In short, the mind ascribes the properties of good and evil to things based on its appetites, even if those things do not really have these properties. And this projection accounts at least partly for the misleading feeling that our actions are determined by our judgments of value.

Though mostly associated with Hume, appeals to the phenomenon of projection were by no means uncommon in the seventeenth century.[21] As theorists of the time understood it, projection is a psychological process by which the mind mistakenly represents objects as having properties that, in fact, merely indicate how those objects affect us. As I will understand it here, a projectivist account of some property has two main features: (i) the false ascription of properties to some object (ii) based on some constitution or affection of our bodies. Examples of projection for seventeenth century theorists paradigmatically included sensory properties such as color, sound, odor, and heat. On projectivist accounts of color, we represent color properties as belonging to objects by virtue of how those objects causally impact the optic nerve. This occurs because we mistake an affection of the body for a property of the object that affects the body.[22]

Spinoza integrates the phenomenon of projection into his account of the mind by making it a constitutive feature of the *imagination*. For Spinoza, the imagination is the first and most common kind of cognition (2p40s2). It is the mind's ability to represent external bodies by representing the way those bodies affect our own bodies: "the affections of the human Body whose ideas present external bodies as present to us, we shall call images of things, even if they do not reproduce the [NS: external] figures of things" (2p17s | G II/106: 465). Thus, for Spinoza, the visual image I form of a tree consists in a mental representation of a state of my body insofar as some part of it is affected by light that has bounced off the tree itself. When I imagine the tree, I represent it *indirectly* in virtue of *directly* representing the changes in my body that the tree has caused.

For Spinoza, the indirect representational structure of the imagination is what makes projection possible. This is because the bodily affections directly represented by ideas of the imagination involve both the nature of external things

[20] LeBuffe (2010: 154-59) argues that Spinoza offers a projectivist theory of value, though he does not fully explain how projection occurs, nor does he draw any connection to the feeling of intelligence at issue here. Spinoza might not be the first to offer a projectivist theory of value. Arguably, Hobbes does so in *Leviathan* (see Darwall 2000). Scribano (2012) makes the case that Spinoza must have read *Leviathan*, and that his engagement with this work accounts for the important shifts in his moral psychology between the *Short Treatise* (completed around 1661) and the *Ethics* (completed around 1675). See Sacksteder 1980 for an account of how much of Hobbes Spinoza might have read.

[21] Hobbes offers a projectivist account of sense properties in the opening passages of *Leviathan* (1651/1668). He was very likely influenced by Galileo's account of heat and of other sense properties in *The Assayer* [*Il Saggiatore*] (1623).

[22] I offer a fuller account of the phenomenon of projection and its place in Spinoza's moral psychology in Moauro unpublished.

and the nature of our own bodies (2p16). Indeed, for Spinoza, the content of these ideas indicates the latter *more* than the former: "the ideas which we have of external bodies indicate the condition of our own body more than the nature of the external bodies. I have explained this by many examples in the Appendix of Part 1" (2p16c2 | G II/104: 463). The images we form of external objects include many properties that are causally explained by features of our bodies. Yet the mind mistakenly represents these properties *as if* they belonged to the external bodies themselves. It projects properties onto external objects based on how those objects causally affect us.

For examples of the phenomenon of projection, Spinoza points us to the Appendix to *Ethics* 1, where he famously offers a sustained criticism of the 'prejudice of divine ends.' This is the belief that God created the world for the sake of human beings, and human beings themselves to worship and glorify him.[23] Spinoza claims that, in the grips of this prejudice, ordinary people form a false conception of God and nature based on their imagination: "[A]ll the notions by which ordinary people are accustomed to explain nature are only modes of imagining, and do not indicate the nature of anything, only the constitution of the imagination. And because they have names, as if they were [notions] of beings existing outside the imagination, I call them beings, not of reason, but of the imagination" (1app | G II/83: 445-46). 'Beings of the imagination' exemplify the two features of projected properties I outlined above. Though they are represented as indicating the natures of things, they in fact indicate "the constitution of the imagination".[24] In other words, they are properties the mind mistakenly attributes to objects based on the condition of our bodies.

Among these beings of the imagination, Spinoza strikingly lists the notions of good and evil: "After men persuaded themselves that everything that happens, happens on their account […] they had to form these notions, by which they explained natural things: *good*, *evil*, *order*, *confusion*, *warm*, *cold*, *beauty*, *ugliness*" (1app | G II/81, 444). Ordinary people "call the nature of a thing good or evil, sound or rotten and corrupt, as they are affected by it. For example, if the motion the nerves receive

[23] "All prejudices I here undertake to expose depend on this one: that men commonly suppose that all natural things act, as men do, on account of an end; indeed, they maintain as certain that God himself directs all things to some certain end, for they say that God has made all things for man, and man that he might worship God" (1app | G II/78, 439-40). Some take this remark as proof that our actions in fact *are* determined by value judgments for Spinoza, since he seems to admit that we act for the sake of ends (see Garrett 1999: 312-14). Indeed, Spinoza later claims that "men act always on account of an end, viz. on account of their advantage, which they want" (1app | G II/78: 440).
I am not persuaded by this reading. Spinoza is likely referring to a *presumption* that human beings act for the sake of ends. We believe (falsely) that we act based on representations of value and assume that all things in nature have similarly been arranged (by God) based on such representations. In fact, it is our appetites that determine all our actions (3p2s). Indeed, Spinoza later provides an account of human ends in terms of appetite: "By the end for the sake of which we do something I understand appetite" (4D7 | G II/210: 547). And appetites, as we've already seen, do not arise from but rather give rise to value judgments (3p9s). So, though Spinoza holds that we necessarily want what has proven to be advantageous or 'useful' to us (cf. 3p12), this is not because we judge it to be good.

[24] Note that the prejudice that God directs all things to ends is not responsible for the phenomenon of projection itself. Regardless of what we think of God, we still perceive objects as if they were colored, warm, etc. What the prejudice *is* responsible for is promoting the belief that these projected properties characterize the natures of the objects.

from objects presented through the eyes is conducive to health, the objects by which it is caused are called beautiful" (1app | G II/82: 445). The example of beauty offers a paradigm instance of projection. We attribute the property of beauty to things based on an affection of our bodies—a motion of the nerves that is "conducive to health." This occurs because the image the mind forms of the beautiful object confuses the constitution of our body with the nature of the object itself. In fact, objects are neither beautiful nor ugly. The notion of beauty indicates not the nature of the object but rather the nature of our own bodies.

In the same way, Spinoza considers goodness a property the mind projects onto things based on how those things affect us. Though we call objects themselves good and evil, these notions in fact indicate something about our own nature. More specifically, they indicate affections of our bodies that determine us to strive for, desire, or have an appetite for an object. This is why Spinoza holds that our value judgments arise from our desires (3p9s). Spinoza's claims about the way we use value terms fit perfectly within a projectivist account. We call objects good or evil because the mind forms an image of them in which the way they affect us—that is, the desire or appetite they bring about—is confused for a property they have in themselves. So, the mind mistakenly attributes value properties to things based on a particular condition of our bodies.[25] For Spinoza, we project value properties onto things based on our appetite for them (3p9s).

5. Diagnosing the Feeling of Intelligence

With this projectivist account of good and evil, Spinoza can explain the feeling of intelligence.[26] For Spinoza, the feeling occurs because projection leads us to invert the order of determination between our desires and our value judgments. I've argued that, for Spinoza, value judgments arise from ideas in which the mind confuses affections of the body with properties of objects themselves. Part of this confusion, I now propose, involves the *causal origin* of our desires. The imagination leads us to experience value properties as the causal determinants of our desires. Since we also experience our desires as determining our actions, we feel that our actions are determined by our judgments of value. This is the feeling of

[25] Thus, Spinoza remarks in the Preface to *Ethics* 4 that the same thing can be called good or bad depending on the affective state of the speaker: "[Good and evil] indicate nothing positive in things, considered in themselves, nor are they anything other than modes of thinking, *or* notions we form because we compare things to one another. For one and the same thing can, at the same time, be good, and bad, and also indifferent. For example, Music is good for one who is Melancholy, bad for one who is mourning, and neither good nor bad to one who is deaf (4pref | G II/208: 545)." A person who is melancholy desires music, so she will call music good. A person in mourning, who is averse to music, will call it evil. Finally, a deaf person, who is not affected by music at all, will call it neither good nor evil.

[26] Some may object to the claim that Spinoza offers a projectivist account of value properties by noting that he provides definitions of good and evil in *Ethics* 4: "By good I shall understand what we certainly know to be useful to us" (4D1 | G II/209: 546). This seems to suggest that Spinoza posits some value properties that are *not* beings of the imagination. Yet the epistemic status of Spinoza's definitions is not always straightforward. Some definitions take as *definienda* imaginative notions, such as the definition of 'contingent': "I call singular things contingent insofar as we find nothing, while we attend only to their essence, which necessarily posits their existence or which necessarily excludes it" (4D3 | G II/209: 546). So, Spinoza's definitions do not show that good and evil are not imaginative notions.

intelligence. So, the phenomenon of projection explains both why we attribute value properties to things and why we take our actions to be motivated by a representation of these properties—and this despite the facts that good and evil are not real properties of things, and that our actions are not determined by our judgments of value.

This explanation of the feeling of intelligence arises within Spinoza's account of *final causes*, which is closely related to his projectivist account of value. For Spinoza, the notion of a final cause indicates an end for the sake of which something occurs. This means that there is a close connection between final causes and intelligent actions. Intelligent actions are determined by judgments of value—judgments that something provides us with a practical reason to pursue it. So, intelligent actions are done for the sake of achieving an end. But Spinoza rejects final causes as a legitimate source of explanation in nature: "[a]ll final causes are nothing but human fictions" because "all things proceed by a certain eternal necessity of nature" (1app | G II/80: 442). For Spinoza, nothing is determined by final causes, and nothing happens for the sake of any end. Consequently, our own actions cannot be determined by final causes.[27]

Despite this, Spinoza recognizes that we frequently appeal to final causes in our explanations of nature and of our own actions. And in the Preface to *Ethics* 4, he offers an account of why this so. He claims that our talk of final causes indicates our appetites, and that we appeal to final causes only because we lack knowledge of what really determines our desires and actions:

> What is called a final cause is nothing but a human appetite insofar as it is considered as a principle, *or* primary cause, of some thing. For example, when we say that habitation was the final cause of this or that house, surely we understand nothing but that a man, because he imagined the conveniences of domestic life, had an appetite to build a house. So habitation, insofar as it is considered as a final cause, is nothing more than this singular appetite. It is really an efficient cause, which is considered as a first cause, because men are commonly ignorant of the causes of their appetites. For as I have often said before, they are conscious of their actions and appetites, but not aware of the causes by which they are determined to want something (4pref | G II/207: 544-45).

Spinoza's account of final causes parallels his projectivist account of value. What we refer to as a 'final cause' is, in fact, an appetite considered a "primary" or first cause of something. When we say that habitation was the final cause of a house, what we mean is that someone built the house *for the sake of* habitation—that is, that the builder's actions were determined by a representation of the value of habitation. By calling habitation a final cause, we imply that the builder's actions were intelligent. But for Spinoza, it is wrong to say that habitation is a final cause, for all final causes are simply human fictions. What calling habitation a

[27] Some scholars, including Sangiacomo (2016), argue that Spinoza's critique of final causes in 1app and 4pref is limited to a particular conception of them prevalent in the Late Scholasticism Spinoza would have encountered in the writings of Franco Burgersdijk and Adriaan Heerebord. See also Garrett 1999. Though there is considerable merit to the claim that the Late Scholastic model of final causation is the one Spinoza targets most directly, I take his arguments to generalize to *all* kinds of final causation. See Viljanen 2011: 123, and Rumbold 2021: 296-97 for more on this point.

final cause in fact indicates is the builder's appetite for housebuilding.[28] Indeed, Spinoza later provides a definition of ends (or final causes) in terms of appetite: "By the end for the sake of which we do something I understand appetite" (4D7 | G II/210: 547).[29] We consider things final causes because we project value onto them based on our appetites. What makes this possible is ignorance about the true causes of our appetites. We can call an object a final cause only if we consider our appetite for it a first cause. But this assumption is false. The builder's appetite is not a first cause of the existence of the house, for like all appetites, it is determined by further causes of which he cannot be aware.

While he eliminates final causes from nature, then, Spinoza provides an explanation of why we nevertheless refer to some things as final causes. This explanation works with his projectivist theory of good and evil to provide an error theoretic account of the feeling of intelligence. The builder treats habitation as a final cause because he is ignorant of the true causes of his appetites. His ignorance of these causes, together with his projection of goodness onto the conveniences of domestic living, leads him to feel that his actions are determined by a judgment that building the house is good. And this feeling contributes to his belief that his actions are freely willed. In fact, the builder's actions are determined by myriad external causes alongside his essential striving to preserve his being, which is insensitive to any judgments of value. His actions are not intelligent, though he inevitably feels that they are.

6. Concluding Remarks

My argument in this paper allows us to fully grasp Spinoza's rejection of free will and his account of the feeling of freedom. His rejection targets more than an incompatibilist conception of free will as an undetermined power of choice. For Spinoza, our actions are neither spontaneous, contingent, nor intelligent. Instead, they are determined by external causes, metaphysically necessary, and not sensitive to any representations of value. Yet due to our necessary ignorance of our action's true causes and to structural features of the imagination, we *feel* that our actions are spontaneous, contingent, and intelligent. Others explain how Spinoza accounts for the feelings of spontaneity and of contingency. Here, I have explained how he accounts for the feeling of intelligence.

[28] The builder develops this appetite "because he imagined the conveniences of domestic life". It might seem from this that the builder *does* act for the sake of an end after all, namely for the convenience of domestic living. Yet the image of convenience can play a role in determining the builder's actions even if it does not involve a value judgment. Indeed, Spinoza claims that the image determines the builder's appetite for house building. But as we saw, Spinoza also claims that our appetites are not determined by value judgments but rather determine them (3p9s). Thus, the while the builder acts with an image of a house in mind, this image does not determine his actions by providing them with a final cause. While the builder acts with an idea of domestic life in mind, he does not act *for the sake of* domestic life—though it inevitably *seems* to him that he does. I thank an anonymous reviewer for prompting me to say more on this point. For more on the role of mental content in Spinoza's theory of action, see Curley 1990, Bennett 1990, Manning 2002, Lin 2006, and Viljanen 2010.

[29] This is another example of a definition that takes as a *definiendum* a notion of the imagination (see note 26). An end for the sake of which we do something is a final cause (as implied by the passage from 4 Preface examined above), and as we've seen, Spinoza considers all final causes human fictions—products of the imagination.

The denial that we satisfy the intelligence condition, along with his error theoretic account of the feeling of intelligence, is crucial to Spinoza's rejection of free will. Leibniz was not wrong to call intelligence the "soul of freedom". The feeling that we act based on our understanding of the good is central to our self-conception as agents—as beings whose behavior is explained in terms of practical reasons and are thus appropriately held responsible for their actions. It is a bold and consequential view of Spinoza's that this self-conception is fundamentally mistaken. We are not endowed with the soul of freedom—with a capacity reasonably taken as central to agency—even if we inevitably feel that we must be.[30]

References

Barry, G. 2016, "Spinoza and the Feeling of Freedom", *Australasian Journal of Philosophy*, 94, 4, 1-15.

Bennett, J. 1984, *A Study of Spinoza's Ethics*, Indianapolis: Hackett.

Bennett, J. 1990, "Spinoza and Teleology: A Reply to Curley", in Curley, E. and Moreau, P.-F. (eds.), *Spinoza: Issues and Directions*, Leiden: Brill, 53-57.

Curley, E. 1990, "On Bennett's Spinoza: The Issue of Teleology", in Curley, E. and Moreau, P.-F. (eds.), *Spinoza: Issues and Directions*, Leiden: Brill, 39-52.

Darwall, S. 2000, "Normativity and Projection in Hobbes' Leviathan", *Philosophical Review*, 109, 3, 313-47.

Deleuze, G. 1988, *Spinoza: Practical Philosophy*, Hurley, R. (trans.), San Francisco: City Light Books.

Descartes, R. 1984 & 1985, *The Philosophical Writings of Descartes*, Vols. 1-2, Cottingham, J., Stoothoff, R., and Murdoch, D. (trans.), Cambridge: Cambridge University Press [CSM].

Descartes, R. 1991, *The Philosophical Writings of Descartes*, Vol. 3, Cottingham, J., Stoothoff, R., Murdoch, D., and Kenny, A. (trans.), Cambridge: Cambridge University Press [CSMK].

Garrett, D. 1999, "Teleology in Spinoza and Early Modern Rationalism", in Gennaro, R. and Hueneman, C. (eds.), *New Essays on the Rationalists*, Oxford: Oxford University Press, 310-35.

Kisner, M.J. 2019, "Spinoza's Activities: Freedom without Independence", in Naaman-Zauderer, N. (ed.), *Freedom, Action and Motivation in Spinoza's Ethics*, New York: Routledge, 37-61.

Klein, J.R. 2003, "Dreaming with Open Eyes: Cartesian Dreams, Spinozan Analyses", *Idealistic Studies*, 33, 2-3, 141-59.

LeBuffe, M. 2010, *From Bondage to Freedom*, Oxford: Oxford University Press.

Leibniz, G. 1985, *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origins of Evil*, Huggard, E.M. (trans.), La Salle: Open Court.

Lin, M. 2006, "Teleology and Human Action in Spinoza", *Philosophical Review*, 115, 3, 317-54.

[30] I would like to thank Donald Rutherford, Ying Liu, and two anonymous referees for valuable feedback on earlier drafts of the paper.

Manning, R. 2002, "Spinoza, Thoughtful Teleology, and the Causal Significance of Content", in Koistinen, O. and Biro, J.I. (eds.), *Spinoza: Metaphysical Themes*, Oxford: Oxford University Press, 182-209.

Melamed, Y.Y. 2017, "The Causes of Our Belief in Free Will: Spinoza on Necessary, 'Innate,' yet False Cognition", in Melamed, Y.Y. (ed.), *Spinoza's Ethics: A Critical Guide*, Cambridge: Cambridge University Press, 121-41

Moauro, L. 2023, "Spinoza's Value Projectivism", Unpublished Manuscript.

de Molina, L. 1988, *On Divine Foreknowledge*, (Part IV of *Concordia*), Freddoso, A. (trans.), Ithaca: Cornell University Press.

Nadler, S. 2006, *Spinoza's Ethics: An Introduction*, Cambridge: Cambridge University Press.

Penner, S. 2013, "Free and Rational: Suárez on the Will", *Archiv für Geschichte der Philosophie*, 95, 1, 1-35.

Ragland, C.P. 2017, "Free Will and Determinism", in Kaufman, D. (ed.) *The Routledge Companion to Seventeenth Century Philosophy*, New York: Routledge, 117-42.

Rumbold, B. 2021, "Spinoza's Analysis of his Imagined Readers' Axiology", *Archiv für Geschichte der Philosophie*, 103, 2, 281-312.

Rutherford, D. 2019, "Deciding What to Do: The Relation of Affect and Reason in Spinoza's *Ethics*", in Naaman-Zauderer, N. (ed.), *Freedom, Action and Motivation in Spinoza's* Ethics, New York: Routledge, 133-51.

Sacksteder, W. 1980, "How Much of Hobbes Might Spinoza Have Read?", *The Southwestern Journal of Philosophy* 11, 2, 25-39.

Sangiacomo, A. 2016, "Aristotle, Heereboord, and the Polemical Target of Spinoza's Critique of Final Causes", *Journal of the History of Philosophy*, 54, 3, 395-420.

Scribano, E. 2012, "La Connaissance du Bien et du Mal: Du *Court Traité* à l'*Éthique*", in Jaquet, C. and Moreau, P.-F. (eds.), *Spinoza Transalpin*, Paris: Éditions de la Sorbonne, 59-78.

Sharp, H. 2007, "The Force of Ideas in Spinoza", *Political Theory*, 35, 6, 732-55.

Spinoza, B. 1925, *Opera*, 4 Vols, Gebhardt, C. (ed.), Heidelberg: Carl Winter [G].

Spinoza, B. 2016, *The Collected Works of Spinoza*, 2 Vols, Curley, E. (trans.), Princeton: Princeton University Press.

Suárez, F. 1597, *Disputationes Metaphysicae*, 2 Vols, Salamanca.

Viljanen, V. 2010, "Causal Efficacy of Representational Content in Spinoza", *History of Philosophy Quarterly*, 27, 1, 17-34.

Viljanen, V. 2011, *Spinoza's Geometry of Power*, Cambridge: Cambridge University Press.

Wee, C. 2006, "Descartes and Leibniz on Human Free-Will and the Ability to Do Otherwise", *Canadian Journal of Philosophy*, 36, 3, 387-414.

Free Will in Leibniz's Thought

Gianfranco Mormino

University of Milan

Abstract

Since the beginning of his activity, Leibniz considers the notion of free will as absurd; he holds this notion not only unnecessary to found moral responsibility but also as an impediment to the correct understanding of divine and human retribution. What prevents many readers to accept this view is Leibniz's insistence on contingency as a requisite of free actions: I argue that the possibility of 'being otherwise' in a different possible world has nothing to do with freedom, which is a perfection, but rather explains the fact that our actions can be wrong.

Keywords: Contingency, Causality, Moral responsibility, Indifference, Incommensurables.

1. The Contextualization

Leibniz starts dealing with the problem of free will in 1671, with a short treatise, now lost, described by himself as a meditation on how human freedom can coexist with divine providence, fate, damnation, God's grace, punishments and rewards (Leibniz 1923ff: II-1, 83). Forty years later, in his *Theodicy*, he will still be dealing with the same issues, but in the light of a life-long enquiry covering all the main subjects of metaphysics. Such relentless interest and constant rework depends on three main reasons:

1) the problem of free will had been at the core of the theological disputes between the rival Christian confessions since the beginning of the 16th century. Deep divisions separated Catholics, Lutherans and Calvinists, and the problem of man's freedom was probably the most controversial one: being for or against free will was an identitarian mark, which drew a precise line between the Reformed and the Catholics. Leibniz, who remained a (very warm) Lutheran throughout his whole life, was deeply engaged in the countless synods, debates and councils which tried to promote ecumenism and his public statements on the problem often oscillated according to the political needs of the moment. Yet, his position never changed: free will, he unfailingly claims, is a philosophical absurdity;

2) on the ethical and juridical level he thinks that the acceptance of free will, far from founding a sound notion of moral responsibility and of legitimate

Argumenta 11, 1 (2025): 51–61
ISSN 2465-2334

DOI 10.14275/2465-2334/20240.mor
First published: 3 June 2024

imputation, would harm it. Quite counterintuitively (but not for those who know Hobbes and Spinoza), Leibniz thinks that guilt and merit are conceivable only within the frame of a strict causality;

3) his commitment to the advancement of science brings him to the acceptance of mechanism in the explanation of physical events. In spite of his effort to create a higher science of dynamics, he never denies that, on the phenomenic level, everything has to—and can—be explained in terms of matter, motion and necessary causality. Whatever goes beyond this realm pertains to a deeper metaphysics but does not imply the existence of free will in any way.

Leibniz uses several different terms referring to *liberum arbitrium*: according to the purpose and to the occasion, they can be more or less pejorative but all of them mean exactly the same thing: *indifferentia pura* (mere indifference), *potestas ad utrumque* (possibility to do either one of two different actions), *potentia agendi, aut non agendi, positis omnibus ad agendum requisitis, scilicet externis* (capability of acting or of non-acting once all external requisites to act are given). All theories admitting mere indifference are equally inacceptable for Leibniz, who compares them to the Epicurean physics; the incomprehensible notion of *clinamen* shows the inconsistency of any position trying to escape a strict notion of causation. Against indeterminism Leibniz argues first of all that God's prescience makes all events (past, present and future) unfailingly determined (Leibniz 2005: 144); secondly, in all true propositions the predicate is contained in the subject, which implies that all contingent futurities are fully determined (Leibniz 2005: 143). Thirdly, he claims that causal chains are permanently and contemporarily effective, so that each state of the world depends on the ones before: "that everything is produced by a strongly determined destiny is as certain as is that three times three is nine. Destiny consists in the fact that all things depend on each other like a chain; therefore everything will unfailingly happen before it happens, as everything unfailingly has happened when it has happened" (Leibniz 1875–1890: 7, 117). Therefore, he criticizes both the Socinians, who limit God's science of future actions (Leibniz 2005: 343), and Aristotle, because of his refusal of the determination of contingent futurities (Leibniz 2005: 324). Leibniz's philosophical major principle (*nihil sine ratione*) rules out any form of hazard, be it the epicurean-lucretian assumption that a body can deviate from its trajectory without any efficient cause, "one of the most impossible of chimeras" (Leibniz 2005: 310–311), or Descartes's attribution of free will to the soul. It is true that our mind can act spontaneously, but it never fails to be harmonized with the perceptions and the movements of the body. Our soul is always determined by impressions, though often imperceptible, which prevail on other ones:

> one might [...] compare the soul with a force which puts forth effort on various sides simultaneously, but which acts only at the spot where action is easiest or there is least resistance. For instance, air if it is compressed too firmly in a glass vessel will break it in order to escape. It puts forth effort at every part, but finally flings itself upon the weakest. Thus do the inclinations of the soul extend over all the goods that present themselves (Leibniz 2005: 322).

Since the universe is infinitely heterogeneous, there is always a difference, even slight, between the alternatives at the moment of the choice. Therefore, the soul always has reasons to incline toward an object more than toward another

one and never finds itself in the absolute indifference which is erroneously depicted in the famous example of Buridan's ass (Leibniz 2005: 150).

In my opinion, it was Hobbes who in particular influenced Leibniz's views on freedom, despite a remarkable difference which we will take into account later in this paper. Most scholars, however, tend to minimize, or even deny, this influence, connecting his views rather to a much wider libertarianism and even assuming his proximity to the Catholic position. In order to tackle this problem, it is necessary to look deeper into Leibniz's arguments.

2. The Ability to Do Otherwise

Since his early writings on liberty and necessity (such as *Confessio philosophi*, 1671–72), Leibniz suggests that the defense of human freedom requires an attenuation of absolute necessity. Such attenuation, nevertheless, is not a refusal of the full determination of all events and can be properly understood only through an analysis of his theory of contingency. At his time, most philosophers thought that only if we admit possibles, something which never was, never is and never will be, we will be able to avoid the danger of absolute necessity, that is Spinozism. Real freedom seems to presuppose the ability 'to do otherwise', which stems from the possibility to *be* otherwise. On this point, in his much later *Theodicy*, Leibniz explains the traditional position in a very clear way: together with understanding and spontaneity, contingency is the third requisite of a sound defense of freedom.

> I have shown that freedom, *according to the definition required in the schools of theology*, consists in intelligence, which involves a clear knowledge of the object of deliberation, in spontaneity, whereby we determine, and in contingency, that is, in the exclusion of logical or metaphysical necessity. Intelligence is, as it were, the soul of freedom, and the rest is as its body and foundation (Leibniz 2005: 303; my italics).

Leibniz, however, is not a follower of the Scholastics and provides his own interpretation of contingency: his hesitation in naming it—"this indifference, this contingency, this non-necessity, if I may venture so to speak" (Leibniz 2005: 310)—is in itself a symptom of his uneasiness about the older views. First of all, he assumes that his notion of contingency "does not prevent one from having stronger inclinations towards the course one chooses; nor does it by any means require that one be absolutely and equally indifferent towards the two opposing courses" (Leibniz 2005: 310).

As well as in the *Theodicy*, in many other late writings he emphasizes the strong connection between freedom and contingency; for example, in his correspondence with Clarke, he distinguishes "liberty, contingency, spontaneity, on the one side, and absolute necessity, chance, coaction, on the other" (Leibniz 1969: 696). The rhetoric thrown against necessitarian positions, which helped him distance himself from all sort of Hobbesianism and Spinozism, actually results into the expression of libertarian views, generally refused in his private writings and hardly compatible with his own metaphysical and moral position. However, for the historian, it is not quite enough to interpret libertarianism as the disguise of a hidden theory, a stratagem conceived as a means to smooth out the religious controversies of his time (and maybe to dodge dangerous accusations). We shall therefore try to seriously take into account his recurring emphasis on the existence of alternatives as a component of human freedom; secondly, we will have to

understand whether such alternatives are implied in any effort to legitimate punishments and rewards, a goal which is indeed at the core of Leibniz's engagement with such issues. As I shall try to prove, Leibniz finds a very innovative and clever way to admit contingency without abandoning a strong determinism.

In the *Theodicy* Leibniz attributes contingency to: a) truths; b) finite substances; c) events; d) actions. It is rather obvious that contingent truths, substances and events are characterised by imperfection and limitation. A contingent truth, for example, cannot be demonstrated in an apodictic way, but only by experience or by showing that it is more reasonable than a different one. It would be quite odd that only for actions contingency should be considered a necessary requirement of freedom, which must be undeniably defined as a perfection. There are therefore good reasons to question whether contingency is seen by Leibniz, just like understanding and spontaneity, as a perfection, and one which must be attributed to God himself like the other two: "the advantage of freedom which is in the creature without doubt exists to an eminent degree in God. That must be understood in so far as it is genuinely an advantage and in so far as it presupposes no imperfection. For to be able to make a mistake and go astray is a disadvantage" (Leibniz 2005: 327). The ability to do otherwise, implied by contingency, is not a perfection; therefore, as I will try to argue, it does not explain how human actions can be free, but rather something quite different, i.e. that human actions can fail. The refusal of free will, whose definition rests on the notion of *indifference*, does not therefore imply the negation of freedom, which is a perfection both in the case of God and of human beings.

Like any other XVIIth-century philosopher, Leibniz cannot avoid considering the problem of free will not only under the viewpoint of human freedom but also under a theological aspect: it is therefore necessary to analyze the problem of whether God's actions, which are by definition free, are contingent. As Robert Adams puts it, "the view that 'God chooses what is best' is contingent must not be regarded as a thesis of Leibniz's philosophy" (Adams 1994: 42).[1] God's fundamental decree, the one which attributes a moral value to the world, is not contingent, as he claims in the *Theodicy*: "one must believe that it is not permitted to do otherwise, since it is not possible to do better" (Leibniz 2005: 197–198). What *is* contingent is the proposition that this world is the best one (Robinet 1955: 393–394). Leibniz oscillates between two different theses about the origin of the decree by which God chooses what is best: the first implies that such choice depends only on his will; the second is derived from his very essence. In the first case he assumes that, within God, there is a 'will to will', which does not imply the contingency of the decree: he simply assumes that the proposition "God wants to choose what is the most perfect" (Leibniz 1923ff: VI, 4B, 1454) is not demonstrable (Leibniz 1923ff: VI, 4B, 1454). The *regressus ad infinitum* of God's volitions, unlike in contingent propositions, does not attain identity through a progressive approximation; rather, it expresses God's spontaneity and his being *causa sui*, while the impossibility to demonstrate a contingent proposition depends on the infinity of the external conditions it implies. God's first decree, then, is the source of all that exists, contingent things included, and not a contingent action itself. In the second explanation, which affirms that God's choice is necessarily determined by his rationality, such necessity does not harm his freedom in any way, since his choice

[1] An opposite opinion in Rescher 1979: 150–152 and Curley 1976: 95–96.

conforms to a perfect rational faculty, which does not admit the ability to do otherwise.

Of course, Leibniz admits that God's choice implies a plurality of possibles: if there were no possibles before the decree, and independent from it, God's act of Creation could not be defined as good. If only *one* world were possible, his action would be the product of an impersonal and axiologically neutral being (as it happens, according to Leibniz, in Descartes's, Hobbes's and Spinoza's thought). Nothing would confer Creation a moral value, which is assured only by the fact that God excludes an infinity of (worse) possible worlds. The question entirely relies on this very point: does the existence of these 'other' possibles (the ones which God *chose* not to create) imply that his action is contingent? And that his freedom is a result of the contingency of his decree? If that were the case, as many believe, then contingency would really be a requisite of freedom even for God himself.

But there is a different option: the contingency we are dealing with pertains neither to the action nor to the agent but, rather, only to the objects of his choice: "the root of God's freedom lies in the possibility of things, i.e. contingency, by which it happens that there are innumerable things which are neither necessary nor impossible, out of which God chooses those who are the most apt to testify his glory" (Leibniz 1948: I, 298). Since God cannot be wrong, the plurality of contingent objects cannot lead him to a different choice, which would be a worse one; his action, therefore, is both free and necessary. The perfect being simply cannot choose otherwise.

The existence of alternatives is not required to justify human moral responsibility as well. First of all, in order to legitimate punishments and rewards, it is not necessary to ascertain whether the agent could have done otherwise; even in the *Theodicy*, Leibniz says that responsibility can coexist with a strong determinism: "the absolute necessity of the sequence of causes would in this matter add nothing to the infallible certainty of a hypothetical necessity" (Leibniz 2005: 160), that is of a necessity depending on external causes. The whole ethical system maintains its whole meaning even if our freedom is not exempt from constriction. There are, according to Leibniz, two kinds of justice: corrective and retributive. The first one implies that the punishments are a cause which produces a change (*amendement*) in the wrongdoer; therefore it is perfectly just to correct a person who could not have acted differently, since the punishment pushes them to act in a better way. In retributive justice (the kind of action which implies a vengeance in order to satisfy the law and the punisher), as well, there is no need to ascertain whether the wrongdoer could have done otherwise, since the goal is the reversal of the advantage gained through the bad action, without any concern for the existence of alternatives.

The fundamental thesis of Leibniz's moral and juridical system asserts that the legitimacy of punishments and rewards depends on their being a necessary means to perfect the world. Contingency, that is the possibility of the opposite, does not play any role in the definition of the minimal requirements of human responsibility, as well as in the attribution of a moral value to God's actions.

On a strictly metaphysical level Leibniz proposes two main solutions to the problem of contingency; however, neither claims that it is a perfection or that it has a role in supporting the existence of free will. The first solution considers contingent things as only hypothetically necessary: they follow from this series of things "i.e. from the harmony of things or from the existence of God" (Leibniz

2003-2005: 12) and, consequently, they are necessary *in the world which God actually created*. It is therefore futile to defend free will by asserting that, in a different world, it would be possible to do otherwise. Leibniz's notion of individuation, and in particular his notion of 'complete concept', implies that the alternative does not concern the same subject, but rather a different one, which exists only in God's mind and which he decided not to create (Leibniz 1875-1890: II, 54).

The second major solution, which he had elaborated ever since *Generales inquisitiones de analysi notionum et veritatum* (1686), claims that contingent propositions cannot be demonstrated because the process of resolving the predicate in the subject is endless (Leibniz 1923ff: VI, 4A, 763). Not even God himself could demonstrate a contingent truth or the existence of a contingent being in a finite number of steps (Leibniz 1923ff: VI, 4B, 1656). There is then no way to derive free will from any of Leibniz's explanations of contingency. Moreover, as we have seen, contingency is neither sufficient nor necessary to impute responsibility to man (Mugnai 1995: 290); responsibility is guaranteed more than enough by the exercise of a conscious will, "wherever it may come from" (Leibniz 1923ff: VI, 1, 542). The foundation of free will on contingency is equally inapplicable to God, who has a perfect intuitive knowledge of the inherence of the predicate to the subject and could not, therefore, make any different connection. If we seek free will inside contingency, the problem cannot be solved. It must therefore be concluded that its very notion is excluded by the logical, ethical and ontological principles of Leibniz's philosophy. It remains then to understand why Leibniz tenaciously tries to give contingency a role in the explanation of man's actions; but, before that, we must present his views about some positions which he deems much more erroneous and dangerous than strict determinism.

3. Pessimistic Theology, Fatalism and Weakness of the Will

The necessity which, in Leibniz's thought, rules human actions must not be confused with the "natural necessity to sin without the succour of divine grace" (Leibniz 2005: 59) that many theologians derive from "the original corruption of the human race, coming from the first sin" (Leibniz 2005: 59). Against the harshest and most pessimistic positions of the Augustinian theology he claims that necessity is compatible with moral improvement; even more, it is necessary in order to understand it. Even Hobbes "[has] shown sufficiently that necessity would not overthrow all the rules of divine or human justice, and would not prevent altogether the exercise of this virtue" (Leibniz 2005: 161). The 'attenuation' of necessity does not imply free will or a looser notion of causation but, rather, the refusal of the idea that Adam's fault has introduced a necessity *to sin*, a position not less dangerous than Spinoza's thought. It is therefore evident that the reasons of his strong opposition to 'absolute necessity' do not lie in a more 'libertarian' view but in the exigency to admit that God acts *sub ratione boni*, not only *sub ratione perfecti*; if it were not the case, punishments would be morally wrong and God could not be distinguished from the tyrannical evil principle of the Manichaeans. Through his theory, Leibniz aims first of all at rejecting the theology of absolute omnipotence, which depicts a God whose decrees do not follow any understandable moral rule. Since Hobbes is often very close to this position, for example in his explanation of Job's book, Leibniz's opposition to his view is perfectly sincere, but does *not* concern determinism.

> And how bitterly did Job expostulate with God, that being just, he should yet be afflicted with so many calamities? God himselfe with open voyce resolved this difficulty in the case of Job, and hath confirmed his Right by arguments drawn not from Jobs sinne, but from his own power [...]. Where wast thou (sayes he) when I laid the foundation of the earth, &c [...]. God by his Right might have made men subject to diseases, and death, although they had never sinned, even as he hath made the other animalls mortall, and sickly, although they cannot sinne (Hobbes 1651: Chapter 15).

A second result of his strategy is the absolute refusal of fatalism: this theory assumes that events happen because 'it is written' (in the stars, maybe) that they must happen. Leibniz claims that our actions are determined by multiple causes, one of which is precisely what we do. As well as Hobbes, he flatly denies that something happens 'no matter what we want': on the contrary, our will is a part of the causal chain which leads to this or that event. He can therefore refuse the 'lazy argument', according to which it is useless to make any effort since everything is already decided; they who behave this way forget that their laziness is exactly one of the causes that leads to their misery. The self-justification of the fatalists is a philosophical mistake and a self-fulfilling prophecy, since it neglects the role they play in their own course of life.

In Leibniz's opinion, the refusal of free will is necessary to explain responsibility: our moral choices are guaranteed by the joined presence of reason and spontaneity. Against the theorists of the inemendable corruption of man, he claims that "the fullest given use of reason" (Leibniz 2003–2005: 17) allows to constrain us 'by things'[2]—definitely not by a mere act of will—to recognize what is good and to push us toward it. Man's responsibility, and consequently the justice of punishments and rewards, is admissible *only* in a deterministic frame: if there were a free will, instead of a necessary connection between causes and effects, men "would not worry themselves about punishments and rewards and would not be led toward good through these means, which would consequently be useless" (Leibniz 1948: II, 483).

Freedom is therefore a conscious use of reason, or *spontaneitas rationalis* (Leibniz 1923ff: VI, 4B, 1380), and presupposes determinism: "the biggest perfection of man consists not merely in that he acts freely but still more in that he acts with reason. Better, these are both the same thing, for the less anyone's use of reason is disturbed by the impulsion of the affections, the freer one is" (Leibniz 1969: 388). They who grant us the absurd privilege of pure indifference, as if we were *entia a se*, detached from the world and free from the regular succession of causes and effects, forget that our actions would be unpredictable, thus making us become anomalies incapable of learning without any possibility to learn from the past and to amend ourselves. On this point Leibniz leans on some of his most important authors: from Lorenzo Valla, Luther and Calvin he draws the notion of a strict predeterminism, from the 'modern' Hobbes, the idea of a world where

[2] "O therefore foolish are we who have scorned the privileges of Nature and God, we demand unknown chimeras and are not contented by the use of reason—the true basis of freedom. Unless it happens from an irrational power we do not think ourselves to be sufficiently free, as if it were not the highest freedom to make use of one's own intellect and will in the most perfect way, and for the intellect to be constrained by things to recognize true goods, and for the will to be constrained by the intellect to embrace them" (Leibniz 2003–2005: 18; my italics).

everything happens because of the necessary effects of God's decree, including not only physical events but also human volitions.[3] We should not then be surprised to find in an early writing a definition of free will which is perfectly adherent to the views of the English philosopher: "he, who can do what he wants and can want what he finds good, has free will enough" (Leibniz 1923ff: VI, 1, 545). Freedom depends on the apparent goodness of the things and on the conditions which present themselves to our will: "to uphold the privilege of free will it is sufficient for us to be placed at a crossroads of life, so that we cannot do something unless we will and we cannot will something unless we believe it to be good" (Leibniz 2003-2005: 17). Will cannot be determined without an *objective* reason, in an erratic way: I can move my left or my right hand, but there are undoubtedly "subtle reasons" (Leibniz 2003-2005: 17) by which I will choose one of them, maybe a physical disposition or, simply, the pleasure to contradict what others predicted.

Medea's famous line "I see and approve the better course, but I follow the worse", when intended as an example of acting against the last judgment of the intellect about the goodness of the choice, is absurd: "she sees the injustice of the deed when she slaughters her own children, but still experiences pleasure from the vengeance as if this were a greater good than the crime was evil" (Leibniz 2003–2005: 18). There is then neither perversion nor weakness of the will (*akrasia*), and the bad actions are as deliberate and causally determined (though wrongly) as the good ones.

Free will is a "manifest illusion [..], a remnant of the delusional philosophy of the past [...], an easy way out that no reasonable man will ever accept" (Leibniz 1923ff: VI, 3, 150–151): there is no 'influence' through which the mind can modify the motion of the body, which follows a conservation rule concerning not only its quantity, as Descartes thinks, but also its direction:

> M. Descartes wished to compromise and to make a part of the body's action dependent upon the soul. He believed in the existence of a rule of Nature to the effect, according to him, that the same quantity of movement is conserved in bodies. He deemed it not possible that the influence of the soul should violate this law of bodies, but he believed that the soul notwithstanding might have power to change the direction of the movements that are made in the body; much as a rider, though giving no force to the horse he mounts, nevertheless controls it by guiding that force in any direction he pleases. But as that is done by means of the bridle, the bit, the spurs and other material aids, it is conceivable how that can be; there are, however, no instruments such as the soul may employ for this result, nothing indeed either in the soul or in the body, that is, either in thought or in the mass, which may serve to explain this change of the one by the other. In a word, that the soul should change the quantity of force and that it should change the line of direction, both these things are equally inexplicable (Leibniz 2005: 156).

4. The Role of Contingency in Leibniz's Account of Freedom

The presence of contingency in Leibniz's theory of *human* actions must be intended quite differently than as a requisite for freedom: the possibility of the

[3] Hobbes's notion of causality, as it is presented for example in the ninth chapter of *De corpore*, is reproposed by Leibniz, under the name of *ratio plena*, in the *Confessio naturae contra atheistas* (Leibniz 1923ff: VI, 1, 490).

opposite is the frame in which our choices are made. It is not a property of the agent but, rather, of the object of the choice. Freedom depends on the spontaneous use of reason and it is a perfection both in God and in us. Contingency aims to explain the difference between real and only apparent reasons, that is the difference between God's and man's actions.[4] By admitting alternatives, Leibniz aims to emphasize the confusion of the context in which the actions of finite beings take place, that is the possibility of mistake. Human actions are always contingent (given the finitenesss of our understanding and the infinite complexity of the world) and sometimes free, when they come from spontaneity and a right understanding. God's decrees, on the contrary, are always free, since his intellect is flawless.

Our reason cannot escape necessity but can fail to understand it, because of the plurality of the objects our intellect has to process: when we are in front of a choice, "our soul is always carried where there is more objective reality. But if those realities are many and minute, it can be a danger for our soul and its happiness. We may call this situation *dissipatio animi*" (Leibniz 1923ff: VI, 4C, 2724). When he claims that sin depends on a lack of attention (Leibniz 1948: I, 365), he thinks of the wrong belief that we could act otherwise, given by the too many alternatives in front of us; dissipation, then, is another aspect of contingency, the most dangerous for our salvation.

Contingency can only be defined in a negative way: contingent truths, as we saw, cannot be demonstrated, because they depend on infinite external conditions. Contingent beings do not have in themselves the reason to exist and are marked by a creatural limitation: "quaecunque igitur veritas analyseos est incapax demonstrarique ex rationibus suis non potest, sed ex sola divina mente rationem ultimam ac certitudinem capit; Necessaria non est. Talesque sunt omnes quas voco Veritates Facti. Atque haec est *radix contingentiae*, nescio an quatenus explicata a quoquam" (Leibniz 1923ff: VI, 4A, 912).

A quite odd way to present this theory is given by Leibniz's analogy with the incommensurables: "in a certain sense, contingent truths are to necessary truths as 'blind' ratios (i.e. of incommensurable numbers) are to the proper and expressible ratios of commensurable numbers" (Leibniz 1923ff: VI, 4B, 1516). In his *Dialogue effectif* Leibniz explains that the impossibility to exactly express such numbers is an imperfection, a participation to Nothingness: "it is evident that this irregularity of incommensurable lines stems from their very essence [...] and cannot be ascribed to God; it is also evident that this incommensurability is an evil that God cannot avoid" (Leibniz 1948: I, 368). Their imperfection does not derive from God's act of creation but predates it: in God's mind there are already contingent beings, whose essence does not depend on His will.

Contingency is a condition halfway between what is determined by itself and the absolutely indetermined, i.e. Nothingness. The freedom of contingent beings is guaranteed by what, in them, comes in a necessary way from God's decree,

[4] "Je voy qu'il y a des gens qui s'imaginent qu'on se determine quelques fois pour le parti le moins chargé, que Dieu choisit quelques fois le moindre bien tout consideré, et que l'homme choisit quelques fois sans sujet et contre toutes ses raisons, dispositions et passions; enfin qu'on choisit quelques fois sans qu'il y ait aucune raison qui determine le choix. Mais c'est ce que je tiens pour faux et absurde, puisque c'est un des plus grands principes du bon sens, que rien n'arrive jamais sans cause ou raison determinante. Ainsi lorsque Dieu choisit, c'est par la raison du meilleur; lorsque l'homme choisit, ce sera le parti qui l'aura frappé le plus" (Leibniz 1875–1890: III, 402).

while their possibility of being mistaken is the effect of their limitation. Contingency, then, is the answer to the great question "Unde malum?". In the 1670's Leibniz uses binary calculus as an analogy to understand the duplicity of finite beings, the 1's being the perfection of what is necessary and the 0's the imperfection of Nothingness. Contingency is the combination of negative and positive; as Fichant says, "la négation est […] l'autre face, insaisissable, de la contingence, et son origine est pour nous aussi inaccessible que celle de la contingence […]. Limitation originale des possibles, mondes possibles qui n'existent pas, tiennent lieu dans la philosophie de Leibniz de la *négativité absente*" (Fichant 1998: 119).

> The question about the origin of error and of evil can be answered only by a theory of contingency: dependence from external conditions, evil, privation of perfection, participation to Nothingness, accidentality are all aspects of the ontological condition of contingent beings (Leibniz 1948: I, 362).

The negative acception of contingency is predictably left in the shadow in his polemics against atheists, materialists and conservative theologians, so much that a reader of the *Theodicy* can easily miss it. Leibniz has very good reasons to insist on the 'coexistence' of freedom and contingency, i.e. on the most libertarian aspects of his philosophy. The cynical way used by Hobbes to describe his own determinism, as well as the harshness of many theologians (both Reformed and Catholics—as for example Malebranche), who identify necessity with the damnation fallen on humanity after Adam's sin, lead Leibniz toward an emphasis on less rigorous positions. Nevertheless, his whole philosophical system, with all his flaws and oscillations, points in a wholly different direction: freedom lies in the correct knowledge of good and in the determination to achieve it, without any real power to change the course of things via a magical power of the soul.

References

Adams, R.M., 1994. *Leibniz: determinist, theist, idealist*. New York: Oxford University Press.

Curley, E.M., 1976, The root of contingency. *In*: H.G. Frankfurt, ed. *Leibniz. A collection of critical essays*. Notre Dame-London: University of Notre Dame Press, 69–97.

Fichant, M., 1998. L'origine de la négation. *In*: *Science et métaphysique dans Descartes et Leibniz*. Paris: Presses Universitaires de France, 85–119.

Hobbes, T., 1651. *De cive*. London: Printed by J.C. for R. Royston, at the Angel in Ivie-Lane.

Leibniz, G.W., 1875–1890. *Die philosophischen schriften von Gottfried Wilhelm Leibniz*, ed. Gerhardt. Berlin: Weidmann.

Leibniz, G.W., 1923ff. *Sämtliche schriften und briefe*. Berlin: herausgegeben von der Preussischen (Deutschen) Akademie der Wissenschaften zu Berlin.

Leibniz, G.W., 1948. *Textes inédits d'après les manuscrits de la Bibliothèque Provinciale de Hanovre*, ed. Grua. Paris: Presses Universitaires de France.

Leibniz, G.W., 1969. *Philosophical papers and letters*, ed. Loemker. Dordrecht: Reidel.

Leibniz, G.W., 2003–2005. *The philosopher's confession,* tr. L. Strickland, http://www.leibniz-translations.com

Leibniz, G.W., 2005. *Theodicy: essays on the goodness of God, the freedom of man and the origin of evil,* tr. E.M. Huggard. LaSalle (Il.): Open Court.

Mugnai, M., 1995. Necessità e contingenza in Leibniz. *In*: A. La Vergata and A. Pagnini, eds. *Storia della filosofia e storia della scienza: saggi in onore di Paolo Rossi*. Scandicci: La Nuova Italia.

Rescher, N., 1979. *Leibniz: an introduction to his philosophy*. Oxford: Blackwell.

Robinet, A., 1955. *Malebranche et Leibniz. Relations personnelles, présentées avec les textes complets des auteurs et de leurs correspondants revus, corrigés et inédits*. Paris: Vrin.

Locke on Free Will and Epistemic Responsibility

Samuel C. Rickless

University of California, San Diego

Abstract

This article summarizes John Locke's considered views on freedom, explaining that freedom is a power of the mind to act in accordance with its volitions, that freedom is a power that can belong only to substances, that we have the freedom to will in many cases, including the power to hold our wills undetermined and thereby suspend the prosecution of our desires. This is a seemingly reasonable account of how our minds work, and should work, when we make (important) decisions. But Locke takes us to be morally responsible and accountable, not just for suspending when it is appropriate, but also for spending our time wisely during suspension, in the proper investigation of what would most conduce to our happiness. The problem is that we are prone to motivated irrationality during suspension when deciding what to investigate and for how long to do so. And thus we need to stop and consider whether we are succumbing to such irrationality before making the ultimate decision. This, I argue, leads to an infinite regress and forces Locke into an unsurmountable dilemma.

Keywords: Locke, Free will, Suspension, Akrasia, Epistemic responsibility.

1. Introduction

In *An Essay Concerning Human Understanding*, John Locke defends a theory of free will and moral responsibility that he thinks accounts for a number of familiar commonsense intuitions, including the fact that human beings sometimes act akratically (i.e., do something bad or wrong, despite believing or knowing that it is bad or wrong) and the fact that human beings are often morally responsible for the choices they make and the acts they commit despite being under the influence of pressing uneasiness(es) pushing them to so choose and act. Although Locke denies that it is literally or strictly speaking true that a human being's will is free, he does think that human beings are often (though not always) free with respect to their acts of willing, and he thinks that there is something that people are referring to, albeit misleadingly, when they use the phrase "free will". This much, I believe, is well understood, though scholars continue to engage in interpretive

Argumenta 11, 1 (2025): 63–78
ISSN 2465-2334

DOI 10.14275/2465-2334/20230.ric
First published: 20 June 2023

debates regarding the details of Locke's views about the nature of, and interconnection between, freedom, volition, and desire.

It is also understood, though perhaps less often advertised, that Locke's theory of free willing is vulnerable to potentially devastating objections. One such objection concerns an infinite regress produced by the interplay between Locke's theory of freedom and his theory of moral responsibility. Locke recognizes that we are continually beset by uneasinesses (i.e., pains at the lack of absent goods) pushing us in one direction or another: on the one hand, I am uneasy inasmuch as I am thirsty; on the other, I see that there is nothing but sugary drink in the fridge, and I am uneasy at the thought of what drinking sugary stuff will do my health; on the one hand, my cousin has dared me to drink what's in the fridge and I am uneasy at the thought of losing face in front of her; on the other hand, my daughters are worried about my health and I am uneasy at the thought of increasing their worries. As Locke sees it, it is possible for me to keep my will undetermined by even the most pressing uneasiness(es) and thereby give myself time and opportunity to examine all relevant considerations and come to a well-reasoned judgment about what it would be best for me to do all things considered. But keeping one's will undetermined is a matter of suspending one's desires, which is itself an act of will. And it would seem that one is morally responsible for one's actions only if one is morally responsible for, because free with respect to, one's act of suspension (or the lack thereof): for if one is not morally at fault for failing to suspend one's desires, it can hardly be the case that one is morally at fault for the actions that result from the prosecution of those desires. But there are various uneasinesses pressing the will with respect to the act of suspension. If one is morally at fault for failing to suspend, this will be only because one is morally at fault for forbearing to suspend the desire to forbear to suspend. Regress therefore beckons: in order to be morally at fault for any particular action, one has to be morally at fault for an infinite number of instances of suspension failure. This seems counterintuitive.[1]

In this article, I would like to raise an additional, and in some ways similar, difficulty for Locke's theory of freedom and moral responsibility. The problem concerns counterintuitive consequences that arise from the interplay between Locke's theory of moral responsibility and the way in which motivated irrationality shapes the epistemic circumstances during episodes of suspension. The worry is that Locke's account of moral responsibility for our decisions and actions can't make sense of our epistemic responsibility during periods of suspension, and that this failure vitiates the account of moral responsibility itself. As far as I am aware, the only way for Locke to surmount this problem is to deny some fairly obvious facts or give up his account of how the will is determined.

2. No Freedom of the Will

I begin with Locke's theory of freedom. The fundamentals of the theory are familiar, but it is necessary to understand the details in order to understand Locke's views on the relation between freedom and moral responsibility.[2]

[1] See, e.g., Rickless 2014: 110-11.

[2] For more on Locke's theory of freedom, see Chappell 1994, Yaffe 2000, 2001, Davidson 2003, Glauser 2003, Lowe 2005, LoLordo 2012, Stuart 2013, Rickless 2000, 2014, 2020, Garrett 2015, Leisinger 2017, and Walsh 2018.

Locke claims that the mind has two basic powers, the power to will and the power to perceive (E 2.21.6).[3] It is very important to Locke that these powers be understood to belong to the mind, and that they not be hypostatized. Although many of Locke's predecessors often wrote as if these powers belong to homuncular faculties, namely the will and the understanding, Locke seeks to disabuse us of this notion. The will and the understanding, according to Locke, are powers, namely the powers to will and to perceive (see E 2.21.5). These powers belong to the mind, and when the mind exercises them, it wills and it perceives. Properly speaking, it is not, therefore, the will that wills or the understanding that perceives: it is the mind (or person) that does the willing and the perceiving.

Freedom, as Locke explains it, is the power to act (that is, to think or to move one's body) according to the direction or determination of one's own mind (E 2.21.8). Thus, if I am able to do A when I will to do A, and I am able not to do A when I will not to do A, then I am free with respect to action A, whether that action be an action of the mind or an action of the body. Thus, as Locke says, if I am able to remove my contemplation from one idea to another when I will to so remove it, then I am free with respect to the act of idea contemplation removal; and if I am able to move my body from one location to another when I will to so move it, then I am free with respect to this particular act of bodily motion (E 2.21.12). "*Liberty*", Locke writes, is, then, "the power a Man has to do or forbear doing any particular Action, according as its doing or forbearance has the actual preference in the Mind, which is the same thing as to say, according as he himself *wills* it" (E 2.21.15).

Locke's seeming equation of willing with preferring in this latter statement of his account of freedom is, as Locke himself recognizes, an overstatement (almost certainly a suboptimal residue of the first edition of the *Essay*).[4] As Locke points out, "it is carefully to be remembred, That *Freedom consists in the dependence of the Existence, or not Existence of any Action, upon our Volition of it, and not in the dependence of any Action, or its contrary, on our preference*" (E 2.21.27). The reason to avoid equating willing with preferring is that the term "preferring" is often used to "signify *Desire* as well as *Volition*", but it is important not to confound the will with desire (E 2.21.30). For it is often the case that "the *Will* and *Desire* run counter", as when someone forces me "to use persuasions to another, which at the same time I am speaking, I may wish [i.e., desire] may not prevail on him" (E 2.21.30). Moreover, although "*Preferring* […] seems perhaps best to express the Act of *Volition*, [it] does it not precisely. For though a Man would preferr flying to walking, yet who can say he ever *wills* it?" (E 2.21.15). As Locke emphasizes, "*desiring* and *willing* are two distinct Acts of the mind" (E 2.21.30), and this is why it is important to Locke that freedom be identified, not with the ability to do what one desires, but rather with the ability to do what one wills: "In this then consists

[3] All references to Locke's *Essay* are to Locke 1975.

[4] Locke revised the chapter on freedom in the *Essay* several times after its initial publication in 1690. The most significant alterations appeared in the second edition (1694), with a few significant changes also appearing in the fourth edition (1700). In the first edition version of E 2.21, Locke had written, for example, that "*Volition* or *Willing*, regarding only what is in our power, *is* nothing but the *preferring* the doing of any thing, to the not doing of it; Action to Rest, *et contra*. Well, but what is this *Preferring*? It *is* nothing but the *being pleased more with the one, than the other*" (E 2.21.28, first edition). I explain below why Locke recognizes this to be a mistake in the second and subsequent editions.

Freedom, (viz.) in our being able to act, or not to act, according as we shall chuse, or *will*" (E 2.21.27).

Freedom, then, is a power. But the will, too, as we have seen, is a power. More particularly, the will is "the *Power* which the mind has, thus to order the consideration of any *Idea*, or the forbearing to consider it; or to prefer the motion of any part of the body to its rest, and *vice versâ* in any particular instance" (E 2.21.5).[5] And the fact that freedom and the will are both powers, Locke explains, makes it impossible for the will to be free in any straightforward or literal sense. "The Question [*Whether Man's Will be free, or no*] is altogether improper; and it is as insignificant to ask, whether Man's *Will* be free, as to ask, whether his Sleep be Swift, or his Vertue square: *Liberty* being as little applicable to the *Will*, as swiftness of Motion is to Sleep, or squareness to Vertue" (E. 2.21.14). The reason why it makes no sense to say that the will is free is, apparently, that "*Liberty*, which is but a power, belongs only to Agents, and cannot be an attribute or modification of the *Will*, which is also but a Power" (E 2.21.14). Powers, then, it seems, cannot belong to powers, but only to agents; and since the will is not an agent, it makes no sense to say that it is free, or, for that matter, unfree.

The claim that powers can belong only to agents is not consistent with other claims Locke makes about powers, and, in truth, I think that Locke does not accept it. In numerous places in the *Essay*, Locke tells us that bodies of various kinds have powers of various kinds. Thus, for example, "Fire has a *power* to melt Gold" and "Gold has a *power* to be melted"; "the Sun has a *power* to blanch Wax, and Wax a *power* to be blanched by the Sun" (E 2.21.1—see also E 2.8.23); the "*Loadstone* [has] the power of drawing Iron" (E 2.23.7); and "Yellowness […] is a Power in Gold, to produce that *Idea* in us by our Eyes, when placed in a due Light" (E 2.23.10). But the sun, wax, magnets, and gold are not *agents*: agents have the power of performing *actions*, but the sun, wax, etc. cannot perform *actions*. This is because actions require "the determination of the *Will*", this being the reason Locke gives for counting forbearances to act as actions (E 2.21.28).

It is not, then, that powers belong only to agents. But something in the vicinity *is* true, and this is that "*Powers* […] *are Attributes only of Substances*" (E 2.21.16). Thus, says Locke, to ask "whether the *Will be free*, is in effect to ask, whether the *Will* be a Substance" (E 2.21.16). And the real problem, then, is not that the will is not an *agent*, but rather that the will is not a *substance*. This is because the will is a power, and "*Powers* are Relations" (E 2.21.19). Locke is here relying on his basic ontology: the world as he understands it is divided into three main categories of entities, namely, substances, modes, and relations. Substances, on this view, are "distinct particular things subsisting by themselves", that is, entities that do not depend for their existence on the existence of other things (E 2.12.6); modes are "Dependences on, or Affections of, Substances" (E 2.12.4); and relations are dependences on at least two things (E 2.25.6). Thus, given the exclusive nature of Locke's categorization, it follows directly from the fact that powers are relations that powers are not substances, and hence cannot themselves possess powers if substances are the only kinds of things that can possess powers. And therefore, if the will is a power, then, given that freedom too is a power, the will cannot (literally) be free.

[5] Again, the use of the word "prefer" in the second clause here is inapposite: the will is not the power to *desire* that one's body move or rest, but rather the power of "commanding the doing or not doing such or such particular action" (E 2.21.5).

There is not a little irony in the fact that the ontological categorization of things into substances, modes, and relations that grounds Locke's claim that there is no such thing as (literal) freedom of the will is something that Locke elsewhere describes as philosophically unhelpful. In a revealing section of the *Essay*, Locke considers the question whether empty space (i.e., space that is devoid of body) is a substance or an accident (i.e., a mode). His answer is that the ideas of substance and accident are neither clear enough nor distinct enough to decide the question (E 2.13.17-20). Indeed, because "we have no *Idea* of what [substance] is, but only a confused obscure one of what it does" (E 2.13.19), Locke expresses thoroughgoing scepticism regarding the usefulness of "the Doctrine of *Substance and Accidents*" when it comes to the "deciding of Questions in Philosophy" (E 2.13.20). And yet, in the chapter on power and freedom, Locke finds himself telling us that the doctrine of substance and relation (which is tightly connected to the doctrine of substance and accident) establishes, in no uncertain terms, that the will cannot be (literally) free. This strikes me as a fairly important philosophical result, indeed one that Locke himself makes much of, expostulating for several sections about the various sillinesses that fall to the lot of those who claim that the will is free because they (as Locke argues, illegitimately) reify the will and treat it as if it were a substance (E 2.21.14-20). And hence the doctrine of substance and accident Locke much maligns leads to consequences for his theory of freedom of will that are more significant than he is elsewhere willing to allow.

3. Free Willing

Having established that there is no such (literal) thing as freedom of the will, Locke turns his attention to the question of whether human beings are free with respect to their *volitions*, i.e., acts of will or, better, individual exercises of the will. Even if my will isn't free, am I not free to will as I please? Locke finds this question ambiguous, and divides it into two separate questions. The first is whether there are any circumstances in which *willing one way or another* is something that one is unable, and hence not free, to avoid. The second is whether there are any circumstances in which *willing to do such-and-such* is something that one is unable, and hence not free, to avoid. There is significant scholarly debate on how Locke answers both of these questions, and I do not propose to repeat the details here. But, relying on past work on the subject, I will now summarize what I take Locke's answers to be, because they set the stage for discussion of one of the central topics of this essay, namely Locke's doctrine of suspension.[6]

Regarding the first question, Locke's answer is that it is sometimes impossible for someone to avoid willing, one way or the other, with respect to a course of action or its opposite. This happens when one is currently engaged in a process and comes to consider, for whatever reason, whether to continue or to stop the process. Locke asks whether "a Man that is walking, to whom it is proposed to give off walking" is "at liberty, whether he *will* determine himself to walk, to give off walking, or no" (E 2.21.24) and his answer to this question is emphatically no. In this sort of case, says Locke, "[t]he Mind has not a power to forbear *willing*", and hence "has not a power to act, or not to act, wherein consists Liberty" (E 2.21.24). If the man is walking and, as he is walking considers whether to continue

[6] For discussions of Locke on the freedom *to* will, see, in particular, Chappell 1994, Rickless 2000, 2014, 2020, and Garrett 2015.

or to stop walking, he cannot avoid willing one way or the other. The reason Locke gives for this is simple: one of the two options, stopping or continuing, "must necessarily follow; and that which does follow, follows by the choice and determination of his Mind, that is, by his *willing* it" (E 2.21.23): thus, if the man continues to walk, it is only because he wills to continue walking, and if the man stops walking, it is only because he wills to stop walking. Either way, whether he continues or stops the process in which he is engaged, the man has no choice but to will. He is, then, in this sort of circumstance, unfree with respect to willing one way or the other.

Regarding the second question, Locke's answer is that it is not merely true but *obvious* that the answer is yes. For remember what freedom is: to be free with respect to an action, whether an action of mind or action of body, is to be able to perform the action if one wills to do so, and to be able not to perform the action if one wills not to do so. Thus, "a Man falling into the Water, (a Bridge breaking under him,) has not liberty, is not a free Agent", for "though he prefers his not falling to falling; yet the forbearance of that Motion [is not] in his Power" (E 2.21.9). Whereas "a Man standing on a cliff, is at liberty to leap twenty yards downwards into the Sea […] because he has the power to leap, or not to leap" (E 2.21.27). And "a close Prisoner, in a Room twenty foot-square, being at the North-side of his Chamber, is at liberty to walk twenty foot Southward, because he can walk, or not walk it: But is not, at the same time, at liberty, to […] walk twenty foot Northward" (E 2.21.27). This account of freedom (and unfreedom) is perfectly general. Consider, now, not whether you are free to A, but whether you are free to *will* to A (where A is some particular action). To be free with respect to that particular act of willing to A is to be able to will to A if one wills to will to A, and also to be able not to will to A if one wills not to will to A. But Locke assumes that willing to will to A is the same as willing to A, and willing not to will to A is the same as not willing to A. And thus, to ask whether one is able to will to A if one wills to will to A is just to ask whether one is able to will to A if one wills to A; and to ask whether one is able to will not to will to A if one wills not to will to A is just to ask whether one is able not to will to A if one doesn't will to A. It is, in other words, to ask "whether a Man can *will*, what he *wills*; or be pleased with what he is pleased with" (E 2.21.25). This is a question that "needs no answer", because the answer to it is obvious, indeed, obviously in the affirmative: *of course* a man can do what he does, *of course* a man can be pleased with what pleases him. Actuality, as a matter of course, entails possibility. Thus, if one wills to A, then one is able to will to A; and if one does not will to A, then one is able not to will to A. Locke concludes that, no matter the circumstances, each one of us is free with respect to any particular act of willing to do such-and-such.

The fact that Locke answers both questions differently explains why he takes pains to distinguish them. Whereas it is always the case that we are free with respect to the act of willing to do such-and-such, it is not always the case that we are free with respect to the act of willing one way or another: indeed, when we are in the middle of engaging in a particular process (e.g., making breakfast, walking to the store, cleaning the dishes, reading a book), we do not have the ability to avoid willing; we must either will that the process continue or that the process cease. As Locke emphasizes, our unfreedom with respect to willing one way or the other occurs in "the far greater number" of circumstances (E 2.21.24). But there are cases in which we are able to avoid exercising our wills, namely,

circumstances in which it is proposed whether to do (or not do) something, initiate or not initiate a course of action, in the near or distant future. Thus, if, while at lunch today, I am considering whether to make coffee or tea at breakfast tomorrow morning, I have the ability *not* to make a decision now about which drink to make tomorrow: that is, I can defer making a decision, one way or the other, and thereby fail to exercise my will. In such a case, I *do* have freedom with respect to willing one way or the other, and this is because the action I am considering is not the continuation or stopping of a process in which I am currently engaged, but rather the initiation and completion of a process that will happen at some future time.

Locke is well aware that, at the time when we consider whether to perform some particular action in the future, we are seldom free "from the sollicitation of our natural or adopted desires" (E 2.21.45). Indeed, whenever we are considering what to do, there are "in us a great many uneasinesses always soliciting, and ready to determine the *will*", and that "the greatest, and most pressing" of these "for the most part" fixes what we will do (E 2.21.47). It is "*uneasiness* alone operates on the *will*" (E 2.21.36) and "determines the *Will* to the successive voluntary actions, whereof the greatest part of our Lives is made up" (E 2.21.33).[7] The uneasiness of desire is a pain "for want of some absent good" (E 2.21.31), and it is only when we feel this kind of want that we are motivated to act (E 2.21.33-39). But Locke insists that, except for unusual circumstances (such as when we face "pains of the Body from want, disease, or outward injuries, as the rack, *etc.* which when present, and violent, operate for the most part forcibly on the *will*"—E 2.21.57), "the mind [has] a power to *suspend* the execution and satisfaction of any of its desires" (E 2.21.47), including even its greatest and most pressing desire, and "keep it from determining the *will*, and engaging us in action" (E 2.21.50). The thesis that we human beings have this power, and that we can exercise it in most cases when we are considering whether or not to perform some particular future action, has come to be known as the doctrine of suspension.

4. Suspension and "Free Will"

What Locke writes about the doctrine of suspension has puzzled commentators, because it suggests that Locke's views on freedom are inconsistent. As I have previously argued, the perceived inconsistency is the result of misinterpretation.[8] Let me see if I can clarify matters further here.

First, what is it to "suspend the execution and satisfaction" of one's desires? Is it an action that follows upon an exercise of the will, or is it a passive occurrence that happens to the mind without its having willed it? The answer is, fairly clearly, that suspension is a *voluntary action*, rather than an unwilled happening. There are several reasons for this.

First, there is the language that Locke elsewhere uses to describe the power to suspend: it is, he says, the power to keep [any particular desire] from determining the *will*", the power of "*standing still*, where we are not sufficiently assured of the way" (E 2.21.50); it is the power to "stop [men's desires] from determining their *wills* to any action, [the power] to hold [their] *wills* undetermined" (E

[7] This is not the view proposed in the first edition of the *Essay*. In E 2.21.29 of the first edition, Locke had written that "*the greater Good is that alone with determines the Will*".
[8] See Rickless 2000, 2014, 2020.

2.21.52). To suspend is to "govern [one's] Passions" and "hinder them from breaking out, and carrying [one] into action" (E 2.21.53); it is to "suspend [i.e., prevent] the act of [one's] choice from being determined for or against the thing proposed" (E 2.21.56). To suspend, he says, is to "hinder blind Precipitancy; […] to stand still" (E 2.21.67). In real life, standing still, stopping, holding, keeping, and hindering all require *effort*. To stand still is to hold one's limbs steady and prevent them from moving. (Effort is not required to prevent the motion of one's body when one is lying down on a comfortable surface, but it is definitely required when *standing*.) To stop or hinder or keep someone from doing something, to hold them back, is to command one's body to act in certain ways, and hence involves volition. The case of the mind, in the matter of preventing one's desires from determining one's will, is not different from the case of the body: this sort of prevention too requires effort and will.

Second, there is one clear textual reference that presupposes that the act of suspension is (or, at least, can be) voluntary. Locke writes: "Nor let any one say, he cannot govern his Passions, nor hinder them from breaking out, and carrying him into action; for what he can do before a Prince, or a great Man, he can do alone, or in the presence of God, *if he will*" (emphasis added—E 2.21.53). What Locke is saying here is that if we are able to suspend the prosecution of our desires in the presence of superiors, then we are able to so suspend when we are alone, if we will to do so. Suspension, then, is the sort of activity that is subject to one's will.

And finally, Locke draws a close connection between one's failure to suspend and one's susceptibility to moral censure and punishment. "[F]rom the not using of [the power of suspension] right", Locke writes, "comes all that variety of mistakes, errors, and *faults* which we run into, in the conduct of our lives, and our endeavours after happiness" (emphasis added—E 2.21.47). In particular, the failure to suspend often results in a "too hasty choice of [one's] own making" and the imposition on oneself of "wrong measures of good and evil", and "the miscarriages that follow on it, must be imputed to [one's] own election". It is for these reasons that "a Man may justly incur punishment, though it be certain that in all the particular actions that he *wills*, he does, and necessarily does will that, which he then judges to be good" (E 2.21.56). Thus, forbearing to suspend when suspension is called for is something for which we may be legitimately blamed or punished. But, as Locke would surely recognize, if the failure to suspend were something that merely happens to us, without being under the control of our wills, then blame and punishment would be inapposite.[9]

The problem that the doctrine of suspension poses for Locke's theory of freedom is that what Locke says about the former appears to contradict the latter. Locke claims, in a variety of passages scattered over several sections of the chapter on power, that liberty "lies in" the power to suspend (E 2.21.47), that this power is "the source of all liberty" (E 2.21.47), "the hinge on which turns the *liberty* of intellectual Beings in their constant endeavours after, and a steady prosecution of true felicity" (E 2.21.52), and "the great inlet, and exercise of all the *liberty* Men have, are capable of, or can be useful to them" (E 2.21.52). These passages suggest that the true liberty of human beings *consists in* the power to suspend, i.e., the power to hold their wills undetermined by their desires, and this appears to contradict Locke's standard and oft-repeated view that liberty (or freedom) consists

[9] For a defense of the contrary view, see Walsh 2014.

in the power to do (and not do) as one wills. Consider, for example, someone who is able to walk if she wills to walk and who is able not to walk if she wills not to walk, but who is, for whatever reason, unable to suspend her desire to walk (or her desire not to walk). According to Locke's official account of the nature of freedom as freedom of action, this person is free; but according to the claim that freedom consists in the power to suspend, she is not free.

If suspending (and, in many cases, the failure to suspend) is the act of holding one's will undetermined, then the ability or power to suspend is the ability or power to hold one's will undetermined. And this, of course, raises the question of whether one has the power to hold (or not hold) one's will undetermined, *if one so wills*. In many cases (not including such extreme circumstances as mental torture), Locke's view is that we can suspend if we will to suspend, and we can fail to suspend if we will not to suspend. According to Locke's overarching definition of freedom, then, we (usually) have freedom with respect to the (mental) act of suspension. The power to suspend (or not suspend) as one wills is therefore a *particular instance or type* of freedom of action, as applied to the mental act of suspension. Locke's view, encapsulated in the somewhat awkward terminology of "source", "hinge", and "inlet", is that this very particular *kind* of freedom is more important than any other. I have the power to lift or not lift my finger if I will, the power to think about philosophy if I will, the power to vote if I will, the power to conjure memories of my childhood if I will. But none of these powers is as critical to my happiness as the power to suspend my desires. For the power to suspend, he says, was "given" to every human being (presumably by God), "[so] that he might examine, and take care of his own Happiness, and look that he were not deceived" (E 2.21.56). The reason is that, as Locke repeatedly says, suspending has a *point* or *purpose*, which is to create the opportunity "to consider the objects of [one's desires]; examine them on all sides, and weigh them with others" (E 2.21.47), in such a way as to "inform [oneself], whether that particular thing, which is then proposed, or desired, lie in the way to [one's] main end, and make a real part of that which is [one's] greatest good" (E 2.21.52). If human beings did not have the power to suspend, they would be at the mercy of their most pressing uneasinesses, very few of which are self-developed or self-created, and, for a variety of reasons relating to the imperfection of their will and judgment, would fall into misery and evil. The fact that we have the freedom to walk and think about philosophy is a happy by-product of the fact that we have the freedom to act more generally; but the reason why we have freedom of action at all, conjectures Locke, is that it makes possible for us the power to suspend, without which we would be miserable. It is for this reason, I think, that Locke tells us that the power to suspend is "the great inlet, and exercise of all the *liberty* Men have, are capable of, or can be useful to them, and that whereon depends the turn of their actions" (E 2.21.52).

There is therefore no inconsistency at the heart of Locke's theory of freedom. Freedom, for Locke, is freedom of action: the ability to do or not do as one wills. Because suspending the prosecution of one's desires is an action, it makes sense to ask whether we are free with respect to that particular type of action. And in most cases, Locke says, we have the freedom to suspend, i.e., the power to suspend or not suspend as we will. As Locke sees it, this freedom to suspend matters more to our felicity than freedom with respect to any other sort of action, because it is guaranteed that, given our various imperfections, we would be miserable and depraved without it. It is for this reason that Locke describes the power of suspension as the "source of all liberty", and as "that, which is (as I think improperly)

call'd *Free will*" (E 2.21.47). As he has already argued, it makes no sense to say that our wills are free, but it *does*, he thinks, make sense to say that we have a particularly important liberty, namely the liberty to hold our wills undetermined, and it is this particular liberty that those who discourse on "free will" and its practical and theoretical importance really (do and should) care about.

5. Akrasia, Moral Responsibility, and Epistemic Duties

I have already said that Locke takes us to be subject to justified blame and punishment if we misuse our power of suspension. How, and under what sorts of circumstances, might this power be misused?

Locke's view of human beings is that they are unmotivated to change their current circumstances unless they find themselves uneasy at the lack of some absent good. Thus, if our needs and wants are met and we are not beset by uneasiness, then we will not lift a finger. But once uneasiness crops up, we are motivated to remove it inasmuch as we judge it to be incompatible with our happiness. This is how we find ourselves looking for food and drink, and then eating and drinking, as a way of removing the uneasinesses of hunger and thirst. Under many ordinary circumstances, there is nothing pathological about any of this activity.

But sometimes we are beset by some pressing uneasiness that threatens to determine our wills to some foolish or evil activity, and it is when facing these sorts of circumstances that it is crucial for us to exercise the power of suspension. The most notable of these forms of practical irrationality is akrasia, or weakness of will, which is when one chooses (and does) something that one knows or believes to be bad, in the belief that it is bad. Locke describes a "Drunkard" who is beset by "*uneasiness* to miss his Companions" and experiences a "habitual thirst after his Cups [that] drives him to the Tavern", even while he recognizes that further inebriation threatens "the loss of health and plenty, and perhaps the joys of another life" (E 2.21.35). On Locke's view, we are justified in blaming (and God would be justified in punishing) the drunkard, not because he experiences a pressing uneasiness to do something bad (to himself or others), but because he forbears to exercise the power to suspend the desire to meet his friends and drink with them at the tavern. Like most people under most circumstances, the drunkard need not immediately choose to pursue the goods at the lack of which he is uneasy. He can hold his will undetermined, and thereby prevent formation of the volition to join his friends at the pub. During this time, says Locke, he has a further power that it is also important for him to exercise if he desires happiness, as all human beings do (E 2.21.54, E 2.21.57, E 2.21.68). This is the power to consider the consequences, the advantages and disadvantages, of (habitually) drinking alcoholic beverages at taverns. If the drunkard exercises the power of suspension without exercising his powers of empirical investigation and ratiocination, he will end up in exactly the position he found himself in at the time he initiated suspension. But if he exercises these latter powers as he should, he will discover just how bad it is for his pocketbook to purchase drinks in taverns and how bad it is for his health to consume excessive amounts of alcohol. Although he might believe, and even know, that it would be better for him to avoid the tavern, he might not be in a position to judge that avoidance of the tavern forms a necessary part of his (long-term) well-being until careful consideration of the evils of paying for and imbibing large quantities of alcoholic drinks lead him to experience an uneasiness that is even more pressing than the original uneasiness at the idea of

missing the pleasure of spending time and drinking with his friends. Indeed, the entire purpose of "examination" or "consideration" during the period of suspension is to "raise desire", a proportionate uneasiness at the thought of the bad results of repeated inebriation and loss of means, sufficient to counteract the initially more pressing uneasiness at the thought of missing the pleasures of the pub (E 2.21.56). And if mere consideration is not sufficient in some cases, then "practice, application, and custom" can contribute to raising countervailing desires "in most" (E 2.21.69).

It should be noted that the drunkard's choice situation, even after he acquires a mountain of information about the effects of alcohol on his metabolism and the price of alcoholic drinks relative to his assets and liabilities, is perhaps not as obvious as Locke makes it out to be (unless it is supposed that, say, a single instance of alcohol consumption is forbidden, say, by God). The drunkard knows that if he goes to the tavern, he is likely to spend a certain amount of money and imbibe a certain amount of alcohol before making his way home. But, unless he is penurious, going to the tavern once will not bankrupt him, and, unless he is right on the edge of contracting cirrhosis or pancreatitis or any of a number of bodily dysfunctions caused by excessive alcohol ingestion, he will not suffer greatly by going to the tavern that evening, except perhaps for a serious headache (and the consequences thereof) the next morning. Even after collating and processing all of the available information that could be relevant to the decision whether to go to the tavern, it should not be obvious to the drunkard that the consequences of going are worse on the whole than the consequences of forbearing to go. This is all consistent with the drunkard recognizing that a course of *repeated visits to the tavern* over a relatively short period of time will likely cause significant financial hardship as well as serious damage to his health. So, perhaps it might have been better for Locke to have chosen a somewhat different example, such as the case of a man who gets a serious thrill from a kind of dangerous activity that has a relatively high probability of leading to his death. Even if he understands how high the probability of death is, his most pressing desire may be to engage in the activity. But if he suspends this desire and then thinks carefully about the probabilities (and perhaps of the consequences of his death for his family and friends), the thinking process during suspension may "raise" a countervailing desire that becomes more pressing and determines his will to forbear from engaging in the activity.

Possession of the power of suspension, as we have seen, is what makes human beings susceptible to moral criticism (and punishment, at least by God) when they misuse it or fail to exercise it when they should. But it should be clear that suspension of one's desires is not, by itself, sufficient to absolve one of moral blame or censure. This is because we have more specifically epistemic duties to discharge during the period of suspension, in addition to the duty to work on ourselves through consideration and practice to raise desires for goods that really are necessary for our happiness: the duty to seek and gather evidence on both sides of the question, the duty to analyze the evidence and thereby determine the probability of such-and-such consequence occurring if so-and-so action is taken and the total amount and intensity of goodness or badness represented by each consequence, and then the duty to form a considered judgment on the basis of this analysis. If one merely suspends without discharging one's epistemic duties during suspension, then one has failed to do as one should and one is subject to just as much blame and censure as would have been appropriate if one had chosen not to suspend to begin with.

The cases of the drunkard and the thrill-seeker may not bring our epistemic duties into high relief, because there is a sense in which both of these individuals already know many things that are relevant to their decisions. So, consider a philosopher, Sally, who recently earned her Ph.D., has applied widely for a job, and is fortunate enough to receive two job offers. One position is at a small, prestigious, but somewhat geographically isolated liberal arts college. The other position is at a large, less prestigious university in a cosmopolitan city that serves as an airline hub. We may imagine that Sally loves both research and teaching, but perhaps enjoys research a bit more than she enjoys teaching. At the same time, the liberal arts college is her alma mater, and she has positive memories of her undergraduate experience there. What should Sally do? Well, she could consult her uneasinesses and go with the one that happens right now to be the most pressing. Perhaps that uneasiness is the thought of missing out on being back in her comfortable undergraduate environment. Or she could hold her will undetermined and do some research. She might discover, for example, that her alma mater made some poor financial decisions that resulted in a diminution of its endowment, and that the faculty who would be her future colleagues are disaffected. And she might discover that the less prestigious university is very well run, that the instructors who would be her future colleagues there are effective and dedicated teachers, and that she would fit seamlessly into the department's plan for future development. At that point, Sally might feel an uneasiness at the thought of missing out on the opportunities that would likely be available at the larger university, an uneasiness that might be greater than her original uneasiness at the thought of missing out on a return to her alma mater. If our philosopher suspends her present desires but doesn't even attempt to find out more about what life would be like at each institution, she would be blameable for failing to gather and analyze the data that would point her to the better alternative.

6. The Problem

So far, what Locke says appears to make a good deal of sense. It appears psychologically respectable as well as morally insightful. Human beings are very occasionally impulsive and sometimes unable to avoid willing (such as when they are on the rack), but, more often than not, they are able to take a step back before making a choice, gather evidence, analyze it, and come up with a more carefully wrought decision. But matters are more complex than Locke recognizes, and the interaction of various factors of which he is aware poses a serious problem for his accounts of freedom, suspension, and moral responsibility. I have already mentioned the problem of infinite regress. But there is another, potentially even more serious problem, in the offing.

Suppose Sally is trying to decide which of the two job offers to accept. Following Locke's strictures, she has not impulsively decided in accordance with her most pressing initial uneasiness (related to her affection for her alma mater) but has exercised her power of suspension and is currently holding her will undetermined. In addition, she has decided to look more carefully at both options and gather evidence that will help her form a more considered judgment, thereby either deepening her former uneasiness or raising a new and more pressing uneasiness to counteract the former. But how much evidence should she gather?

Sally has, let us say, a limited amount of time in which to make her decision. If she waits too long, she will forgo both job opportunities, each of which clearly

ranks higher than her current situation. But, of the time she has available, she could spend more or less time gathering relevant information. Suppose she discovers, by reading the Chronicle of Higher Education, that the endowment of her alma mater (call it "AM") has suffered a loss that is well short of catastrophic. A mutual friend has told her that the instructors in the philosophy department at AM are disaffected, but then, in discussions with each of them, she discovers that there are reasonable explanations for their current disaffection, and that there are reasons to think that the disaffection will be short-lived. There are too many instructors at the larger university (call it "LU") for Sally to talk with them all. The conversations she has had with them have been pleasant but relatively short, and, for understandable reasons, none of them has been particularly forthcoming about personal or professional challenges. Should Sally continue gathering information, or stop?

Suppose Sally is really attached to AM. She remembers her time there with great fondness. During her years there as an undergraduate, the students were well-supported and the instructors decently remunerated. The location is geographically isolated, but there were compensating advantages, such as opportunities for communing with nature and making close friends in a small town. So, Sally stops her investigation and now consults her levels of uneasiness at the thought of not joining AM and at the thought of not joining LU, allows her current most pressing uneasiness to determine her will, and decides to accept the offer from AM. According to Locke, it appears that Sally has done her duty: counteracting her initial desires, she postponed making a decision, gathered relevant information, analyzed it, and only then allowed her levels of uneasiness to fix her choice.

And yet all, we may suppose, is not so rosy. The question whether to gather more evidence can only be answered by an act of will. There is, we may reasonably suppose, uneasiness at the thought of gathering more evidence, but also uneasiness at the thought of failing to gather more evidence. It is possible for the decision whether or not to gather more evidence to be determined by the most pressing current uneasiness or to be postponed by suspending the execution and satisfaction of one's present desires. Perhaps the degree of Sally's attachment to AM is influencing the current balance of uneasiness at the thought of gathering more information. Locke himself recognizes that "*when we compare present Pleasure or Pain with future*, […] *we often make wrong Judgments of them*", because "[o]bjects, near our view, are apt to be thought greater, than those of a larger size, that are more remote" (E 2.21.63). So, perhaps Sally's judgment that she has gathered enough information is being distorted by her attachment to AM, in ways of which she is not currently aware, just as judgments of the relative size of present pleasures and future pains are unconsciously distorted by "*the weak and narrow Constitution of our Minds*" (E 2.21.64).

It seems, then, that, faced with the question of whether to stop gathering information, Sally should suspend her desire to answer that question in order to consider whether this desire is being influenced in potentially distorting ways by her attachment to AM. If she doesn't suspend the desire to decide whether to stop gathering information and simply allows her present balance of uneasinesses to determine her will with respect to the question, then she will stop gathering information, thereby putting her in a position of deciding whether to accept AM's offer on what may well be an inadequate evidentiary basis. So, on the one hand, by suspending her desires and gathering evidence relevant to deciding which job

offer to accept before making her decision on the basis of the evidence, Sally has discharged her duty. But, on the other hand, by failing to consider whether her decision to stop gathering evidence is subject to motivated irrationality, Sally has done wrong.

One might think that Sally has a way out, one that requires only the smallest of tweaks to Locke's conception of what is required to fulfill one's epistemic duties during the time that one's desires are suspended. Before deciding whether to stop gathering evidence, Sally can suspend her desires on this matter in order to determine whether it is best, on the whole, for her to stop gathering evidence. During *this* period of suspension, Sally can investigate whether her desire to stop gathering evidence is being unduly influenced by her attachment to AM. If the result of that investigation is that she has been unduly influenced, then she should decide to counteract her attachment and continue gathering evidence relevant to deciding which job offer to accept. But if the result of that investigation is that she has *not* been unduly influenced, then she should stop gathering evidence and decide which job offer to accept on the basis of the evidence she has already gathered.

Unfortunately, matters are not so simple. Suppose Sally is now investigating whether her attachment to AM is distorting her judgment about whether to stop gathering evidence regarding which job offer to accept. That investigation, too, requires the gathering of evidence. For example, Sally could ask herself whether strong attachments of various kinds have had an irrational influence on some of her past decisions. She could talk to her friends and relatives about whether they think of her attachment to AM as excessive and unjustified. She could consult with her therapist about whether she has overly romanticized past relationships with students and faculty at AM during her undergraduate years. Suppose Sally has gathered some evidence (on the question of whether her attachment to AM is distorting her judgment about whether to stop gathering evidence relevant to the decision of which job offer to accept), but she recognizes that she could also gather more. What should she do? Keep gathering evidence, or stop?

And now, I hope, we can see the nature of the problem. The decision to continue or stop gathering evidence about the effects of her AM-attachment on her decision to stop or continue gathering evidence relevant to another decision will be the result of an act of will that is ordinarily determined by the balance of uneasinesses concerning the options between which she is deciding. But perhaps the most pressing of *those* uneasinesses is the result of some factor that threatens to have a distorting influence on her judgment. So, Sally might be well-advised to hold her will undetermined at this point, in order to gather evidence that is relevant to the question of whether some emotional factor is irrationally affecting her desire to stop gathering evidence about the effects of her attachment to AM on her desire to stop gathering evidence about which job offer to accept. But there is no principled end to this process. It would seem that, in order to discharge her epistemic duties, Sally must continue investigating, at an unending series of levels, whether her decision to stop or continue gathering evidence regarding some decision or other is being unduly influenced by some psychological factor of which she may be currently unaware. But this is impossible. Sally must, in fact, stop gathering evidence at some point and make a decision. And yet the decision to stop investigating, at whatever level it occurs, is going to be unavoidably arbitrary and subject to criticism.

The result of all this is that Locke is caught in a dilemma. On the one hand, he can say that once Sally has suspended at the first level and conducted some sort of investigation, no matter how curtailed, in advance of making her decision, then she has done all that could reasonably be expected of her. But this is counterintuitive. On the other hand, he can say that, once Sally has suspended at the first level, she should gather evidence up to a point that is determined by an investigation into whether she should stop or continue gathering evidence, which involves gathering evidence up to a point that is determined by an investigation into whether she should stop or continue gathering evidence, and so on, *ad infinitum*. But this, of course, is impossible.

The problem, as I see it, is generalizable to all circumstances in which decisions are called for in life, except perhaps the most trivial. The nub of it is that the discharging of our duties regarding the decisions we need to make is not limited to the suspension of desire. If suspension of desire were all that was required to discharge our decisional duties, then the problem would be avoided: when faced with a decision one should suspend desire and use reason to determine which decision would be best, at which point one should cease suspending and make the decision. Locke often writes as if matters were this simple. He describes the process of examination during suspension as "*consulting a guide*", and "the determination of the *will* upon enquiry" as "*following the direction of that Guide*" (E 2.21.50). But, of course, there is no *literal* guide, and there is no manual for how to make decisions in life. Occasionally, Locke suggests that the matter may be more complex. Thus, he tells us that human beings should suspend their desires "till they have duly and fairly *examin'd* the good and evil of it, as far forth as the weight of the thing requires". "This", he writes, "we are able to do; and when we have done it, we have done our duty, and all that is in our power; and indeed all that needs" (E 2.21.52). But, of course, there is no manual or objective measuring stick by which to determine whether one's examination has been "due" or "fair" or what level of investigation is proportional to the "weight of the thing". The most, perhaps, that can be said is that, other things equal, matters of less moment require less, whereas matters of greater moment require more, in the way of investigation. But that, by itself, does not diminish or erase the problem that I have attempted to elucidate.

Apart from the fact that our duties extend not merely to the suspension of desire but also to the gathering and analysis of evidence, there is also the fact that *every* decision, including the decision whether to stop or continue gathering evidence, results from the balance of various uneasinesses that could be influenced by irrational emotions or beliefs, unless the desires fed by those uneasinesses are suspended. Thus, the only way for Locke to escape the problem I have described would be to (i) deny that our duties extend past suspension (and perhaps some degree of examination during suspension), (ii) insist that there is a simple or objective way of determining the appropriate extent of examination during suspension, or (iii) deny that the decision whether to continue or stop examination is usually determined by the most pressing uneasiness, unless the desire corresponding to that uneasiness is suspended. But (i) involves a significant moral mistake, (ii) involves a significant factual error, and (iii) involves giving up a central plank of Locke's (second edition) theory of how the will is determined. To my mind, Locke is well and truly caught.

References

Chappell, V. 1994, "Locke on the Freedom of the Will", in Rogers, G.A.J. (ed.), *Locke's Philosophy: Content and Context*, Oxford: Oxford University Press, 101-21.

Davidson, J.D. 2003, "Locke's Finely Spun Liberty", *Canadian Journal of Philosophy*, 33, 203-27.

Garrett, D. 2015, "Liberty and Suspension in Locke's Theory of the Will", in Stuart, M. (ed.), *A Companion to Locke*, Oxford: Blackwell, 260-78.

Glauser, R. 2003, "Thinking and Willing in Locke's Theory of Human Freedom", *Dialogue*, 42, 695-724.

Leisinger, M.A. 2017, "Locke's Arguments Against the Freedom to Will", *British Journal for the History of Philosophy*, 25, 642-62.

Locke, J. 1975, *An Essay Concerning Human Understanding*, Nidditch, P.H. (ed.), Oxford: Clarendon Press.

LoLordo, A. 2012, *Locke's Moral Man*, Oxford: Oxford University Press.

Lowe, E.J. 2005, *Locke*, London: Routledge.

Rickless, S.C. 2000, "Locke on the Freedom to Will", *Locke Newsletter* (now *Locke Studies*), 31, 43-67.

Rickless, S.C. 2014, *Locke*, Oxford: Wiley-Blackwell.

Rickless, S.C. 2020, "Locke on Freedom", *The Stanford Encyclopedia of Philosophy*, Zalta, E.N. (ed.), https://plato.stanford.edu/archives/spr2020/entries/locke-freedom/

Stuart, M. 2013, *Locke's Metaphysics*, Oxford: Oxford University Press.

Walsh, J. 2014, "Locke and the Power to Suspend Desire", *Locke Studies*, 14, 121-57.

Walsh, J. 2018, "Locke's Last Word on Freedom: Correspondence with Limborch", *Res Philosophica*, 95, 637-61.

Yaffe, G. 2000, *Liberty Worth the Name: Locke on Free Agency*, Princeton: Princeton University Press.

Yaffe, G. 2001, "Locke on Refraining, Suspending, and the Freedom to Will", *History of Philosophy Quarterly*, 18, 373-91.

Hume on Free Will

Lorenzo Greco

University of L'Aquila

Abstract

In this essay, I discuss David Hume's reasoning on free will as he presents it in *A Treatise of Human Nature* and *An Enquiry concerning Human Understanding*. I proceed by showing how Hume's compatibilist solution acquires meaning in the light of his sentimentally based science of human nature, which conceives human beings as reasonable, social, and active creatures. Within Hume's empiricist, naturalistic, and sceptical approach, we deal only with perceptions and never with things themselves, and human experience is structured in a causal order which allows us to organise both the way we experience the world and our existence in relation to that of others. In such a scenario, the question of free will depends on human practices, such as the attribution of responsibility, which follow a causal order and are not affected by metaphysical doubts about the loss of responsibility if determinism were true. I argue that Hume traces responsibility back to the expression of feelings for or against particular characters; people become the object of judgements of responsibility in so far as, through their actions, they show that they possess characters of a certain kind which reflect a whole series of dispositions and traits, empirically verifiable and causally explainable, acquired over time. I conclude by highlighting how free will may represent a problem on a practical level once moral or religious issues come into play and why this is not so for Hume.

Keywords: Hume, Free will, Responsibility, Human nature, Common life.

1. Introduction

In this essay, I offer an exegesis of David Hume's notion of free will. I take the relevant sections of *A Treatise of Human Nature* and *An Enquiry concerning Human Understanding*—that is, T 2.3.1-2 and EHU 8[1]—and go through the arguments there presented to exemplify Hume's position and how he defends it. Although

[1] I shall quote *A Treatise of Human Nature* in the body of the text as T, followed by the book, part, section, and paragraph numbers, and the page of the Selby-Bigge/Nidditch edition (SBN). I shall quote *An Enquiry concerning Human Understanding* in the body of the text as well as EHU, followed by the section and paragraph numbers, and the page of the Selby-Bigge/Nidditch edition.

Argumenta 11, 1 (2025): 79–92
ISSN 2465-2334

DOI 10.14275/2465-2334/20230.gre
First published: 17 July 2023

my intentions are primarily reconstructive, I shall suggest, as a possible interpretation, that his assertions on free will can only be fully appreciated against the background of his ambition to elaborate a "science of MAN" (T Intro 4; SBN xv).

First, it is necessary to elucidate the science of human nature that Hume is thinking about. To judge from what Hume claims in the first book of the *Treatise* ("Of the Intellect") and the first *Enquiry*, this science consists in an experimental, empirical, and *a posteriori* approach aimed at accounting for the powers and limits of the human mind. In doing so, Hume follows the "way of ideas" introduced by René Descartes and developed by John Locke and George Berkeley. Hume works with perceptions, which he distinguishes into impressions and ideas; they provide the material that, together with associative principles such as resemblance, contiguity, and causality, constitute the human mind. However, the science of human nature turns out to be much broader than the associationism with which Hume wants to account for the human mind. Hume, in fact, is also a naturalist, i.e., he sees human beings as creatures endowed with a body and guided by sentiments and passions, and he devotes the two successive books of the *Treatise*, "Of the Passions" and "Of Morals", to the study of human beings thus conceived. This study continues in his subsequent works, from the *Enquiries concerning Human Understanding and concerning the Principles of Morals* to *A Dissertation on the Passions* to the numerous *Essays, Moral and Political*, and then the monumental *History of England*. What emerges from looking at Hume's work taken as a whole is that the objects of his science, human beings, are conceived as flesh and blood creatures who live and act in a natural world. However, Hume's appeal to the natural world must be correctly understood: his naturalism places human beings within a practical dimension, considering them as agents "in the common course of the world" where they can be observed "in company, in affairs, and in their pleasures" (T Intro 10; SBN xix). Within this practical context, human beings reveal themselves as "reasonable" beings, but also as "sociable" beings as well as "active" beings (EHU 1.6; SBN 8). Therefore, from the way Hume sets the issue, human beings are not reducible only to their minds, nor are they only bodies whose functions can be explained biologically; they are first and foremost social creatures whose conduct is appropriately illustrated by looking at the interactions they have with each other. Hume's conclusions on free will remain unclear unless one keeps this background in mind.

2. Uniform Perceptions

Let us see in detail how Hume proceeds. In the *Treatise*, Hume introduces his reasoning about free will as part of a broader representation of the sentiments and passions that characterise human nature. He starts from the relation of the will with direct passions. These, in Hume's taxonomy, arise immediately from good and evil, pleasure and pain; passions such as desire and aversion, grief and joy, hope and fear are of this type. As in the case of the passions, the will, for him, is only an impression of which we become aware the moment we express it: "[B]y the *will*, I mean nothing but *the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind*" (T 2.3.1.2; SBN 399). If it is permissible to speak of the will, for Hume, this is all we can say about it: we have knowledge only of perceptions, which in turn are distinguished into impressions and ideas; therefore, for the empiricist Hume the will, manifesting itself as an impression, is to all intents and purposes

nothing more than an impression, which cannot be further defined. It is from this observation that Hume proceeds to account for the problem of the relationship between freedom and necessity and, with it, the problem of free will.[2]

Hume sets the question through a parallel between the necessity existing among external bodies and that which exists in the mind. Notice that for him, in both cases, any consideration we can make concerns those perceptions of which the will also consists. For both matter and the human mind, we cannot go beyond the observation that we always and only deal with perceptions, never with things in themselves: "[I]n no single instance the ultimate connexion of any object is discoverable, either by our senses or reason, and […] we can never penetrate so far into the essence and construction of bodies, as to perceive the principle, on which their mutual influence depends" (T 2.3.1.4; SBN 400). Hume is here showing the sceptical aspect of his philosophical perspective. Since the ultimate essence of things and the principles that determine the connection between things are inevitably obscure to us, Hume wonders where our idea of necessity could possibly come from. In his view, it is the result of the "constant union" we find among objects—that is, among the perceptions we have of them—that leads our mind to infer, following "an uniform and regular conjunction" (T 2.3.1.4; SBN 400), that particular effects will always necessarily follow from particular causes. Therefore, if it makes sense to speak of necessity, it is nothing more than the result of the concurrence of these two elements: a constant conjunction of perceptions and an inference of the mind. It is not a question of ascertaining the existence of a principle of connection between objects—given Hume's *a posteriori* approach, nothing of the sort can be known—but of accounting for an operation of the mind that projects necessity onto things whenever it faces cases of constant union between our perceptions.

Hume's point is that this type of reasoning applies to the judgements we make about both the external world and human conduct: in either case, the process of ascribing necessity is the same. For him, human conduct expresses the same regularities of external bodies, that is, human conduct, too, involves a constant union of causes (people's motives) and effects (people's actions): "Whether we consider mankind according to the differences of sexes, ages, governments, conditions, or methods of education; the same uniformity and regular operation of natural principles are discernible. Like causes still produce like effects; in the same manner as in the mutual action of the elements and powers of nature" (T 2.3.1.5; SBN 401). In a famous passage in the first *Enquiry*, Hume declares that human conduct—as directly observed and as witnessed in history—exhibits homogeneity, which applies as much to individuals as to society as a whole. It is worth quoting it at length:

> It is universally acknowledged, that there is a great uniformity among the actions of men, in all nations and ages, and that human nature remains still the same, in its principles and operations. The same motives always produce the same actions: The same events follow from the same causes. Ambition, avarice, self-love, vanity, friendship, generosity, public spirit; these passions, mixed in various degrees, and

[2] Here I confine myself to examining the will in so far as it is relevant to the question of free will. For a more specific discussion, and criticism, of the notion of the will in Hume, see Connolly 1987; Keutner 1987; Stalley 1986. For a reconstruction of the scholarly debate on Hume's notion of free will, see Millican 2011.

> distributed through society, have been, from the beginning of the world, and still are, the source of all the actions and enterprizes, which have ever been observed among mankind. Would you know the sentiments, inclinations, and course of life of the Greeks and Romans? Study well the temper and actions of the French and English: You cannot be much mistaken in transferring to the former *most* of the observations, which you have made with regard to the latter. Mankind are so much the same, in all times and places, that history informs us of nothing new or strange in this particular. Its chief use is only to discover the constant and universal principles of human nature, by shewing men in all varieties of circumstances and situations, and furnishing us with materials, from which we may form our observations, and become acquainted with the regular springs of human action and behaviour (EHU 8.7; SBN 83).

This homogeneity enables us to explain why people acted in given ways and to make predictions about future conduct. Human actions may show inconstancy and uncertainty now and then; however, when this happens, there must necessarily be a cause unknown to us which, if we were aware of it, would account for the unusual behaviour. The mechanism that allows us to judge the actions of human beings is always the same:

> When any phænomena are constantly and invariably conjoin'd together, they acquire such a connexion in the imagination, that it passes from one to the other, without any doubt or hesitation. But below this there are many inferior degrees of evidence and probability, nor does one single contrariety of experiment entirely destroy all our reasoning. The mind ballances the contrary experiments, and deducting the inferior from the superior, proceeds with that degree of assurance or evidence, which remains (T 2.3.1.12; SBN 403).

There is a whole series of behaviours that seem to confirm the unpredictability of human conduct but which, on the contrary, are explained and justified because one cannot help but look at people's demeanour as causally determined. That is the case with spontaneous actions or those done in haste or in an unpremeditated manner. Why do we admit them, and why do we not condemn those who do them? Precisely "because a hasty temper, tho' a constant cause in the mind, operates only by intervals, and infects not the whole character" (T 2.3.2.7; SBN 412). Or, Hume continues, take the case of repentance: we accept it when it corresponds to a change in the character of the one who is repenting. A criminal is such when his conduct is moved by "criminal passions or principles in the mind" (T 2.3.2.7; SBN 412; see EHU 8.28-30; SBN 97-99). If these criminal principles are removed, our judgement of him changes. So, what matters to Hume when it comes to describing human behaviour is the uniformity of human conduct, by which we can identify, through the observation of human actions, the principles of human nature as manifested in individual characters. Consistent actions allow us to trace the motives behind them; these motives, in turn, function as an interpretative lens through which to look at actions and explain their meaning:

> Hence likewise the benefit of that experience, acquired by long life and a variety of business and company, in order to instruct us in the principles of human nature, and regulate our future conduct, as well as speculation. By means of this guide, we mount up to the knowledge of men's inclinations and motives, from

> their actions, expressions, and even gestures; and again, descend to the interpretation of their actions from our knowledge of their motives and inclinations. The general observations, treasured up by a course of experience, give us the clue of human nature, and teach us to unravel all its intricacies (EHU 8.9; SBN 84-5).

Human beings may well behave in ways that are not always predictable, but this is not to deny that their conduct is regulated in terms of cause and effect:

> [W]e know, in general, that the characters of men are, to a certain degree, inconstant and irregular. This is, in a manner, the constant character of human nature; though it be applicable, in a more particular manner, to some persons who have no fixed rule for their conduct, but proceed in a continued course of caprice and inconstancy. The internal principles and motives may operate in a uniform manner, notwithstanding these seeming irregularities; in the same manner as the winds, rain, clouds, and other variations of the weather are supposed to be governed by steady principles; though not easily discoverable by human sagacity and enquiry (EHU 8.15; SBN 88).

We can admit exceptions in individual behaviour because we reason in terms of uniformity. These exceptions are the result of "conceal'd" causes that are "contrary" (T 2.3.1.12; SBN 404) to the normal behaviour we expect of individuals. And yet, they are still causes that produce effects which can only be explained if we admit the doctrine of necessity: "From the observation of several parallel instances, philosophers form a maxim that the connexion between all causes and effects is equally necessary, and that its seeming uncertainty in some instances proceeds from the secret opposition of contrary causes" (EHU 8.13; SBN 87).

As in the case of external bodies, for human conduct, too, the constant union of causes and effects—i.e., motives and actions—leads the human mind to infer the existence of one from the other. Again, it is the human mind that organises experience according to a causal arrangement conceived as necessary, a necessity that, for external bodies as well as for human things, has no other reason for being than the ascertainment of regularities and hence the inference that is drawn from them by the mind. It is thereby that "the force of *moral evidence*" can be affirmed, which "is nothing but a conclusion concerning the actions of men, deriv'd from the consideration of their motives, temper and situation" (T 2.3.1.15; SBN 404). A uniformly structural experience of the world enables human beings to move with confidence in it and relate to one another in the certainty that they can predict people's actions according to repeatedly observed relationships: "A prince, who imposes a tax upon his subjects, expects their compliance. A general, who conducts an army, makes account of a certain degree of courage. A merchant looks for fidelity and skill in his factor or supercargo. A man, who gives orders for his dinner, doubts not of the obedience of his servants" (T 2.3.1.15; SBN 405). Hume's explanation of causal necessity reveals itself most clearly in a practical context of this kind, where social creatures such as humans regulate their actions on the basis of what others do. Causal necessity makes it possible to manage how human beings interact with each other thanks to the fact that they, given their natural constitution, cannot but infer that specific actions will follow from specific behaviours because the constant experience of particular behavioural dynamics confirms this.

3. A Feeling of Freedom

It is worth insisting that when Hume speaks of necessity, he is not expressing a judgement about things *per se*, since it is here that the key to understanding what Hume means by free will lies. To the extent that it is correct to claim that Hume is a determinist,[3] one must keep in mind that his determinism must be seen in the light of the mental mechanisms whereby human beings order experience according to causal necessity. Humean determinism is not a statement regarding discernible relations between things in themselves: "[T]he *necessary connexion* is not discover'd by a conclusion of the understanding, but is merely a perception of the mind" (T 2.3.1.16; SBN 405-406). Hume reiterates this in the first *Enquiry*, as well, when he defines the notion of necessity by comparing it to that of cause. Hume recalls that the latter can be defined in two ways, either by referring to the constant conjunction between objects or to the inference that "the mind" or "the thought" establishes between one object and another (see T 1.3.14.31; SBN 170; EHU 7.29; SBN 76-77). Necessity, too, can be understood in these two senses, either by conceiving two objects as constantly conjoined together or by looking at "the inference of the understanding from one object to another", and Hume specifies that these two senses, "indeed, are at bottom the same" (EHU 8.27; SBN 97). That is why he confidently concludes that natural evidence and moral evidence are the same in nature: "And, indeed, when we consider how aptly *natural* and *moral* evidence cement together, and form only one chain of argument betwixt them, we shall make no scruple to allow, that they are of the same nature, and deriv'd from the same principles" (T 2.3.1.17; SBN 406). In both cases, these are ways of structuring human experience from a content that is always the same: perceptions. And in both cases, it is not a question of grasping the ultimate structure of the world but of accounting for how sentiment-driven creatures such as human beings bring coherence to their experience, give meaning to their existence in common, and thus survive.[4]

That being so, Hume embraces a *compatibilist* solution, i.e., he believes that human beings can be said to be free even if determinism is true.[5] At first glance, it may seem that reconciling human freedom and determinism is a contradiction in terms; indeed, much of the debate on free will revolves around this very theoretical juncture, with "incompatibilists" arguing that the truth of determinism *ipso facto* denies that human beings can be said to be free. However, if the specific way in which Hume understands necessity applies, this is not the case. In fact, "[t]he necessity of any action, whether of matter or of the mind, is not properly a quality in the agent, but in any thinking or intelligent being, who may consider the action, and consists in the determination of his thought to infer its existence from some preceding objects" (T 2.3.2.2; SBN 408). Hume is thinking of a necessity thus defined, which allows us to account for people's actions by tracing them back to their characters. If we could not do this, we would not evaluate

[3] "It is universally allowed that nothing exists without a cause of its existence, and that chance, when strictly examined, is a mere negative word, and means not any real power which has anywhere a being in nature" (EHU 8.23; SBN 95).

[4] In this regard, his determinism differs from that of Thomas Hobbes, who instead conceives determinism as concerning things in themselves. See Chappell 1999.

[5] For a debate on Hume's compatibilism, compare Beebee & Mele 2002; Harris 2005: Chpt. 3; 2012; Millican 2010; Penelhum 2000; Pitson 2016; Russell 2021a.

anyone's actions since we can only identify individuals as distinguished by their specific characters. If necessity fails, it is no longer possible to understand what it means for human beings to be free; if necessity is absent, so are causes, and human action without causes corresponds to chance. But this, Hume observes, goes against our experience of human action, which is always causally explicable. So, to be free does not consist in acting according to what Hume calls "liberty of *indifference*"—a liberty that is defined regardless of any causal determination—but in acting without being forced or prevented; that is, in Hume's terminology, in acting according to "liberty of *spontaniety*" (T 2.3.2.1; SBN 407; the typo is in the Humean text).

Hume can hence offer a definition of "liberty" as "*a power of acting or not acting, according to the determinations of the will;* that is, if we chuse to remain at rest, we may; if we chuse to move, we also may" (EHU 8.23; SBN 95).[6] Although Hume speaks of a "power" on our part to act as we please, this power must nonetheless be comprehended in line with a freedom understood as spontaneity. It so happens that human beings, when they perform any action, do not feel that it is the result of any necessity but experience it as depending on nothing, that is, "there is a *false sensation or experience* even of liberty of indifference; which is regarded as an argument for its real existence" (T 2.3.2.2; SBN 408). The point, Hume continues, is that we feel that our actions, unless explicitly guided by some external force, are the fruit of our will. Consequently, we also feel that our will is not compelled by anything. Nevertheless, given the universe of perceptions with which human beings deal, the will itself—i.e., for Hume, the feeling we have of it—is nothing more than a perception. Although this is the sensation we experience in the first person, yet, from a third-person point of view, our actions can always be explained by tracing them back to previous motives or circumstances: "We may imagine we feel a liberty within ourselves; but a spectator can commonly infer our actions from our motives and character; and even where he cannot, he concludes in general, that he might, were he perfectly acquainted with every circumstance of our situation and temper, and the most secret springs of our complexion and disposition" (T 2.3.2.2; SBN 408-9; see EHU 8.22 fn 1; SBN 94). If, then, we can be said to be free because we have the power to act or not to act, this power, for Hume, corresponds to our subjective feeling of possessing a will that does not depend on anything. Free will is nothing more than our sensation of being free; yet, this subjective "posture of the mind"—to refer to the title of a book by a renowned Humean, Annette Baier (Baier 1985)—can be explained, and always is, in terms of an observable cause—a motive of ours—which has given rise to our action. To affirm the existence of a will devoid of any causal determination goes beyond what we can claim about ourselves and the external world; as we have seen, for Hume, the essence of things is precluded from us, so it is simply nonsensical to claim both that the will is es-

[6] The paragraph continues as follows: "Now this hypothetical liberty is universally allowed to belong to every one, who is not a prisoner and in chains. Here then is no subject of dispute". Despite Hume's speaking here of a "hypothetical liberty" and not of liberty as spontaneity, as he did in the *Treatise,* his project of defending a notion of freedom understood as spontaneity, fully reconcilable with necessity, remains unchanged in the two works. See Bricke 1996: 235-36; 2008: 207-209; Garrett 1997: 119-21; 2015: 198; Penelhum 2000: 161; Pitson 2006: 223; 2016: 381-82; Russell 1995: 12-13; 2015: 231-32; Stroud 1977: 145; Wright 2009: 170.

sentially free and that it is essentially determinate, if by "determinate" we mean a sense of necessity beyond our perceptions.

The concern of those who claim that, if everything is determined, then we can never be said to be truly free, does not hold true for Hume. That is because, for him, the level of analysis from which such an assertion is made proves misplaced. That kind of reasoning has a semblance of meaning only within a philosophical context in which causal necessity is understood as something that concerns the essence of things. But for the philosopher, like Hume, who has embraced "true philosophy" (T 1.4.3.9; SBN 222) and understood that "philosophical decisions are nothing but reflections of common life, methodized and corrected" (EHU 12.25; SBN 162), conceiving of causal necessity in this way is meaningless:

> To reconcile the indifference and contingency of human actions with prescience; or to defend absolute decrees, and yet free the Deity from being the author of sin, has been found hitherto to exceed all the power of philosophy. Happy, if she be thence sensible of her temerity, when she pries into these sublime mysteries; and leaving a scene so full of obscurities and perplexities, return, with suitable modesty, to her true and proper province, the examination of common life; where she will find difficulties enough to employ her enquiries, without launching into so boundless an ocean of doubt, uncertainty, and contradiction! (EHU 8.36; SBN 103).

The true scientists of human nature, those who are able to correctly illustrate how human nature works, are for Hume not the metaphysicians but the political and moral philosophers, that is, those who carefully observe human beings in action in order to derive the general principles of their conduct: "These records of wars, intrigues, factions, and revolutions, are so many collections of experiments, by which the politician or moral philosopher fixes the principles of his science; in the same manner as the physician or natural philosopher becomes acquainted with the nature of plants, minerals, and other external objects, by the experiments, which he forms concerning them" (EHU 8.7; SBN 83-84). It is only within areas such as "history […] politics […] morals [and] criticism" (EHU 8.18; SBN 90) that the dialectics between free will and necessity can be correctly understood.

4. Being Responsible

So, how does Hume harmonise determinism with free will in such a context? On the one hand, he accepts the definition of liberty of spontaneity and rejects that of liberty of indifference. On the other hand, the determinism that Hume admits never goes back as far as grasping the necessary links between things in themselves. Hume's determinism stops at the level of perceptions and involves how humans react to constant events; the metaphysical arguments about the reconcilability of free will with the truth of determinism, which occupy so much of the debate on the subject, do not apply to him because they are formulated on a level on which he does not, and does not wish to, place himself. Hume is interested in accounting for the ways in which human beings make sense of both their experience of the world and the way they relate to each other, and human beings happen to function by virtue of causally organised reasoning as much in the physical world as in the moral world. Given all this, freedom of the will—defined in the terms admitted by an understanding of liberty of spontaneity—is

therefore admissible. Indeed, Hume's real problem turns out to be not so much that of reconciling determinism and free will as that of accounting for how human beings can exercise liberty of spontaneity given the passions and emotions that characterise human nature and given certain human practices that depend on this sentimentally marked nature of theirs. In other words, Hume reverses the order in which we generally proceed when addressing the question of free will: he takes as established certain human practices which are confirmed by our experience of human conduct and only then does he turn back to address the issues of human liberty, free will, and determinism.

These practices provide the normative criteria for understanding in what sense human beings can be said to be free. Of course, if human nature were different—that is, if it revealed different sentimental regularities from those we experience and manifested itself through different practices—then perhaps things would also be different with regard to what we can say about free will. But, given how we have experienced human nature so far, this is not the case, and we are just as ready to regard as a liar one who, after a long journey, tells us that he has met human beings with passions totally different from ours, as we are to discard the account of a historian who ascribes to past peoples a human nature irreconcilable with that known to us:

> Should a traveller, returning from a far country, bring us an account of men, wholly different from any, with whom we were ever acquainted; men, who were entirely divested of avarice, ambition, or revenge; who knew no pleasure but friendship, generosity, and public spirit; we should immediately, from these circumstances, detect the falsehood, and prove him a liar, with the same certainty as if he had stuffed his narration with stories of centaurs and dragons, miracles and prodigies. And if we would explode any forgery in history, we cannot make use of a more convincing argument, than to prove, that the actions, ascribed to any person, are directly contrary to the course of nature, and that no human motives, in such circumstances, could ever induce him to such a conduct (EHU 8.8; SBN 84).

At the same time, human conduct and, with it, the question of free will, can only be fully made sense of by looking at people's actions as interrelated: "The mutual dependence of men is so great in all societies that scarce any human action is entirely complete in itself, or is performed without some reference to the actions of others, which are requisite to make it answer fully the intention of the agent" (EHU 8.17; SBN 89). This theoretical framework is reminiscent of what would be argued two centuries later by Peter Strawson, who based the discussion of the relationship between determinism and free will on the observation that people relate to each other by virtue of certain "reactive attitudes", such as resentment and attribution of responsibility, which are embedded in human nature, defining its perimeter of meaning. These reactive attitudes are not affected by the truth of a causal necessity that makes all human actions determined (Strawson 1962).[7] Hume's discussion of free will also pivots on the notion of responsibility. Hume observes that the real problem of causal necessity, when applied to human affairs, turns out to be fundamentally both moral and religious

[7] Paul Russell systematically juxtaposes Hume and Peter Strawson on free will. See Russell 1995, 2017, 2021a, 2021b.

in nature. If the doctrine of necessity were true, it is often complained, then human responsibility would disappear, and, with it, one of the pillars of ethics and religion; consequently, the doctrine of necessity must necessarily be false. This reasoning takes for granted what has yet to be proved—that we enjoy a freedom to act otherwise that does not depend on any external causal necessity—and does so because of a worry which, although understandable, does not by itself shed any light on how things actually are: "When any opinion leads us into absurdities, 'tis certainly false; but 'tis not certain any opinion is false, because 'tis of dangerous consequence" (T 2.3.2.3; SBN 409). On the contrary, Hume argues that only the presence of a causal necessity between our motives and actions explains what we actually mean when we say that someone is responsible for something, thus providing the strongest guarantee for morality and religion. The very notions of merit and demerit require the principle of necessity to apply; how is it possible to assess whether the actions that someone performs are to their credit or detriment if we cannot trace them back to their character?

> The constant and universal object of hatred or anger is a person or creature endow'd with thought and consciousness; and when any criminal or injurious actions excite that passion, 'tis only by their relation to the person or connexion with him. But according to the doctrine of liberty or chance, this connexion is reduc'd to nothing, nor are men more accountable for those actions, which are design'd and premeditated, than for such as are the most casual and accidental (T 2.3.2.6; SBN 411).

Therefore, for us to understand merit and demerit, we must necessarily observe human conduct as structured according to a causal dynamic. To deny this would be to adopt an interpretative lens that would eventually lose sight of its object: people as they are in "common life" and, with them, their characters and conduct. For this reason, according to Hume, even those who deny necessity cannot help but resort to it, whether they like it or not: "But so inconsistent are men with themselves, that tho' they often assert, that necessity utterly destroys all merit and demerit either towards mankind or superior powers, yet they continue still to reason upon these very principles of necessity in all their judgments concerning this matter" (T 2.3.2.7; SBN 411-12). In this sense, on closer inspection the question of free will turns out to be a problem of terminology only (see EHU 8.1; SBN 80-81).

Note how the doctrine of necessity also plays into Hume's hands in order to criticise religious conceptions of morality. If, in fact, one wants to establish a link between morality and religion and conceives the divinity as the supreme legislator, its laws will be ineffective, and no obedience will be due to them without a necessary connection between cause and effect in human actions. Yet, if we grant that everything is determined according to a continuous chain of causes and effects and we place the divinity at the origin according to the doctrine whereby "[t]he ultimate Author of all our volitions is the Creator of the world" (EHU 8.32; SBN 99), then we must also admit that the divinity is responsible for everything human beings do. If this is the case, then "[h]uman actions […] either can have no moral turpitude at all, as proceeding from so good a cause; or if they have any turpitude, they must involve our Creator in the same guilt, while he is acknowledged to be their ultimate cause and author" (EHU

8.32; SBN 100). That, of course, creates serious problems for religious moralists and divines:

> [I]f human actions can be traced up, by a necessary chain, to the Deity, they can never be criminal; on account of the infinite perfection of that Being from whom they are derived, and who can intend nothing but what is altogether good and laudable. Or […] if they be criminal, we must retract the attribute of perfection, which we ascribe to the Deity, and must acknowledge him to be the ultimate author of guilt and moral turpitude in all his creatures (EHU 8.33; SBN 100-101).

This dead-end does not affect Hume, who believes in necessity but not in the divine origin of the world.[8] His way of interpreting necessity in human affairs is sufficient to support the convictions that moral beauty and ugliness "are founded in the natural sentiments of the human mind" (EHU 8.35; SBN 103) and that these can be explained within the context of "common life" in which the abstract reasoning of both metaphysical and religious thinkers—but more generally "any philosophical theory or speculation whatsoever" (EHU 8.35; SBN 103)—fail to catch on:

> These enlarged views may, for a moment, please the imagination of a speculative man, who is placed in ease and security; but neither can they dwell with constancy on his mind, even though undisturbed by the emotions of pain or passion; much less can they maintain their ground, when attacked by such powerful antagonists. The affections take a narrower and more natural survey of their object; and by an economy, more suitable to the infirmity of human minds, regard alone the beings around us, and are actuated by such events as appear good or ill to the private system (EHU 8.34; SBN 101-2).

5. Concluding Remarks

In this essay, I have discussed Hume's reasoning on free will as he presents it in the *Treatise* and the first *Enquiry*. I have argued that his reflections cannot be understood without taking into account his project of elaborating a science of human nature. Hume's compatibilist solution acquires meaning in the light of a sentimentally understood human nature that fully reveals itself within the perspective of "common life", in which human beings are conceived as reasonable, social, and active creatures. I have observed that, for Hume, basing the discourse on free will on metaphysical considerations about the truth or falsity of determinism is misplaced. Within Hume's empiricist, naturalistic, and sceptical approach, we deal only with perceptions and never with things themselves. Given the way human nature works, experience is structured in a causal order which allows us to organise both the way we experience the world and our existence in relation to that of others; in turn, the question of free will depends on human practices, such as the attribution of responsibility, which follow a causal order and are not affected by the philosophers' metaphysical doubts about the loss of responsibility if determinism were true. On his part, Hume traces respon-

[8] According to Russell, the insurmountable problems Hume finds for religion in his discussion of free will are further evidence of the underlying irreligiousness of Hume's overall philosophical project. See Russell 2021a, 2021b: essays 11-13. Hume's irreligion is discussed at length in Russell 2008.

sibility back to the expression of certain feelings for or against particular characters; people become the object of judgements of responsibility in so far as, through their actions, they show that they possess characters of a certain kind. Characters reflect a whole series of dispositions and traits, empirically verifiable and causally explainable, acquired in the course of time, whereby people can be identified. In such a context, free will may represent a problem mainly on a practical level once moral or religious issues come into play.

And indeed, free will becomes a problem only for those who fear that the truth of determinism deprives human beings of such a morally fundamental practice as the attribution of responsibility—a concern voiced by the so-called "libertarians". According to them, compatibilist solutions such as Hume's—but also Peter Strawson's appeal to "reactive attitudes"—fail to guarantee what really matters, namely, that people's actions truly derive from their will and not external factors. However, we have seen that, for Hume, such a concern reveals a fundamental confusion. Our feeling free does not justify the belief that our will is not subject to any causal dependence. From what we can observe, it would seem instead that our will consists precisely of those contingent attributes and qualities which many libertarians reject as not expressing our deepest practical identity as autonomous agents (see Campbell 1951). On closer inspection, their worry corresponds to the need to think of ourselves as "self-made selves" (Russell 1995, 130),[9] i.e., as individuals who are the sole authors of their own choices, expressing "self-determination" (Strawson 1994), or "agent causation" (O'Connor 1995). This worry, for Hume, finds expression only from within our moral practices and does not bring into play any metaphysical reflection on the status of free will. The only legitimate expression of free will is in terms of freedom as spontaneity, not freedom as self-determination and even less freedom as chance.

The libertarian stance reveals a definite moral conviction according to which the perimeter of what is ethically relevant stops where voluntary agency ends. Here again, Hume's position is more nuanced: what is ethically relevant goes beyond intentional action, referring to a "constitutive luck" (Williams 1981; see Russell 1995: Chpt. 9) that occupies a primary role in determining both what we are and what we can be held responsible for. Although one may continue to insist that the sphere of ethics must enjoy a status of its own that is prior to moral luck and impervious to it (see Nagel 1979), for Hume, this is not enough to make it so. On the contrary, it seems that what can be the object of praise or blame—and, with it, of other fundamental ethical concepts, among which that of responsibility—is recognisable only *a posteriori* and therefore is irremediably exposed to the influence of chance. (Think, in this sense, of the lack of a precise boundary between virtues and talents, on the one hand, and vices and defects, on the other, which Hume speaks of in T 3.3.4-5 and *An Enquiry concerning the Principles of Morals*, App 4.) By linking ethics to voluntariness, libertarians end up conceiving of responsibility in all-or-nothing terms: either we are fully responsible for what we committed, i.e., for what we freely wanted to do, or we cannot say that we caused a given outcome for which we can be either praised or blamed. However, if one takes moral luck seriously, it no longer seems possible to describe responsibility as a phenomenon that either exists or does not exist but as something that involves degrees, whereby one can be more

[9] In turn, Russell takes this expression from Dennett 1984, Chpt. 4.

or less responsible depending on the circumstances. An approach like Hume's acknowledges this since it understands responsibility as springing from the character of the agents, offering an explanation that does not need to refer to metaphysical or religious presuppositions. Also, this approach does justice to the fact that ethics represents an aspect of human nature, and that human nature can unfold in many ways depending on many different factors, not all necessarily under our control. People are regarded as more or less responsible on the basis of the development of their character, which in turn is subject to other influences, such as circumstances and upbringing.[10]

References

Baier, A. 1985, *Postures of the Mind: Essays on Mind and Morals*, Minneapolis: University of Minnesota Press.

Beebee, H. & Mele, A. 2002, "Humean Compatibilism", *Mind*, 442, 111, 201-23.

Bricke, J. 1996, *Mind and Morality: An Examination of Hume's Moral Psychology*, Oxford: Clarendon Press.

Bricke, J. 2008, "Hume on Liberty and Necessity", in Radcliffe, E.S. (ed.), *A Companion to Hume*, Oxford: Blackwell, 201-16.

Campbell, C.A. 1951, "Is 'Freewill' a Pseudo-problem?", *Mind*, 240, 60, 441-65.

Chappell, V. (ed.) 1999, *Hobbes and Bramhall on Liberty and Necessity*, Cambridge: Cambridge University Press.

Connolly, J. 1987, "David Hume and the Concept of Volition: The Will as Impression", *Hume Studies*, 13, 2, 276-305.

Dennett, D. 1984, *Elbow Room: The Varieties of Free Will Worth Wanting*, Oxford: Clarendon Press.

Garrett, D. 1997, *Cognition and Commitment in Hume's Philosophy*, Oxford-New York: Oxford University Press.

Garrett, D. 2015, *Hume*, Abingdon-New York: Routledge.

Harris, J. 2005, *Of Liberty and Necessity: The Free Will Debate in Eighteenth Century British Philosophy*, Oxford-New York: Oxford University Press.

Harris, J. 2012, "Free Will", in Bailey, A. & O'Brien, D. (eds.), *The Continuum Companion to Hume*, London-New York: Continuum, 214-26.

Hume, D. 1975, *Enquiries concerning Human Understanding and concerning the Principles of Morals*, Selby-Bigge, L.A. & Nidditch, P.H. (eds.), Oxford: Clarendon Press.

Hume, D. 1978, *A Treatise of Human Nature*, Selby-Bigge, L.A. & Nidditch, P.H. (eds.), Oxford: Clarendon Press.

Keutner, T. 1987, "David Hume and the Concept of Volition: The Will as Wish", *Hume Studies*, 13, 2, 306-22.

[10] I would like to thank Roger Crisp, Dan O'Brien, and two blind referees for their helpful comments. This essay was written with the contribution of the Grant Agency of the Czech Republic for the research project "Virtues, Old and New: Virtue Ethics in Hume and Mandeville" [GAČR 20-02972S] carried out in cooperation with the University of Hradec Králové.

Millican, P. 2011, "Hume, Causal Realism, and Free Will", in Allen, K. & Stoneham, T. (eds.), *Causation and Modern Philosophy*, Abingdon-New York: Routledge, 123-65.

Millican, P. 2010, "Hume's Determinism", *Canadian Journal of Philosophy*, 40, 4, 611-42.

Nagel, T. 1979, "Moral Luck", in *Mortal Questions*, Cambridge: Cambridge University Press, 24-38.

O'Connor, T. 1995. "Agent Causation", in O'Connor, T. (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, Oxford-New York: Oxford University Press, 61-79.

Penelhum, T. 2000. "Hume and the Freedom of the Will", in *Themes in Hume: The Self, the Will, Religion*, Oxford: Clarendon Press, 156-76.

Pitson, A.E. 2006. "Liberty, Necessity, and the Will", in *The Blackwell Guide to Hume's Treatise*, Traiger, S. (ed.). Oxford: Blackwell, 216-31.

Pitson, A.E. 2016, "Hume, Free Will, and Moral Responsibility", in Russell, P. (ed.), *The Oxford Handbook of Hume*, Oxford-New York: Oxford University Press, 380-400.

Russell, P. 1995, *Freedom and Moral Sentiments*, Oxford-New York: Oxford University Press.

Russell, P. 2008, *The Riddle of Hume's Treatise: Skepticism, Naturalism, and Irreligion*, Oxford-New York: Oxford University Press.

Russell, P. 2015, "'Hume's Lengthy Digression': Free Will in the *Treatise*", in *The Cambridge Companion to Hume's* Treatise, Ainslie, D.C. & Butler, A. (eds.), Cambridge: Cambridge University Press, 230-51.

Russell, P. 2017, *The Limits of Free Will: Selected Essays*, Oxford-New York: Oxford University Press.

Russell, P. 2021a, "Hume on Free Will", in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/fall2021/entries/hume-freewill/.

Russell, P. 2021b, *Recasting Hume and Early Modern Philosophy: Selected Essays*, Oxford-New York: Oxford University Press.

Stalley, R.F. 1986, "The Will in Hume's *Treatise*", *Journal of the History of Philosophy*, 24, 1, 41-53.

Strawson, G. 1994, "The Impossibility of Moral Responsibility", *Philosophical Studies*, 75, 1-2, 5-24.

Strawson, P.F. 1962, "Freedom and Resentment", *Proceedings of the British Academy*, 48,187-211.

Stroud, B. 1977, *Hume*, London-New York: Routledge.

Williams, B. 1981, "Moral Luck", in *Moral Luck: Philosophical Papers 1973-1980*, Cambridge: Cambridge University Press, 20-39.

Wright, J.P. 2009, *Hume's "A Treatise of Human Nature": An Introduction*, Cambridge: Cambridge University Press.

Kant on Free Will

Derk Pereboom

Cornell University

Abstract

For Kant transcendental freedom consists in the power of agents to produce actions without being causally determined by antecedent conditions in exercising this power. He contends that we cannot establish whether we are actually or even possibly free in this sense. Kant claims only that our conception of ourselves as transcendentally free involves no inconsistency, and that as a result the belief that we are free in this sense meets a relevant standard of minimal credibility. Justification of this belief ultimately depends on practical reasons: the need to believe that we are subject to moral obligation and that we are morally responsible. I argue that the belief that we are transcendentally free does satisfy an appropriate standard of minimal credibility, but that the practical reasons Kant adduces for it should be controversial.

Keywords: Kant, Free will, Transcendental freedom, Determinism, Moral obligation, Moral responsibility.

1. Introduction

Immanuel Kant (1724–1804), dissatisfied with the sort of compatibilism advocated by philosophers such as David Hume (1711–1776), the hard determinism of Spinoza (1632–1677), and any view on which nature is indeterministic, advocated a bold conception of freedom on which human agents are the undetermined sources of their actions while at the same time all events in the natural world are thoroughly causally determined. In another respect, Kant's theory is cautious: he maintains that our being free in this sense—transcendentally free—cannot be established theoretically, that is, from evidence available to us. Rather, our evidence can establish only that believing we are transcendentally free involves no inconsistency, and our justification for this belief must instead rely on practical reasons.[1]

[1] Spinoza presents his position on free will in his *Ethics* (Spinoza 1985), and Hume develops his view in the *Treatise of Human Nature* (Hume 1978) and in *An Enquiry Concerning Human Understanding* (Hume 2000). An earlier version of my statement of Kant's view on free will can be found in Pereboom 2006.

Argumenta 11, 1 (2025): 93–110
ISSN 2465-2334

DOI 10.14275/2465-2334/202521.per
First published: 30 November 2025

2. The Terms of the Debate

It is best to begin with some definitions of key terms. The term 'free will' as it is used in the philosophical debate has several distinct senses, and the answer may depend on which sense is meant. In a first sense, to have free will is to have alternative possibilities for choice and action:

> **Free will AP** (for "alternative possibilities"): free will is an agent's ability, at a given time, either to act or to refrain: that is, if an agent acts with free will, then she instead could have refrained at that time from acting as she did.

A second sense, which predominates in the contemporary discussion, links free will to moral responsibility:

> **Free will MR** (for "moral responsibility"): free will is an agent's ability to exercise the control in acting required to be morally responsible for an action.

In the history of the free will debate, causal determinism of some sort has been taken to be the main threat to our having free will in each of these two senses:

> **Causal determinism**: every event has causal antecedents that render it inevitable.

The parties to the free will debate are traditionally grouped into camps with reference to whether causal determinism and free will are compatible:

> **Compatibilism**: our having free will is compatible with causal determinism, more specifically, with all of our actions being causally determined by factors beyond our control.
>
> **Incompatibilism**: our having free will is not compatible with causal determinism, with all of our actions being causally determined by factors beyond our control.

Compatibilists in Kant's philosophical ambit include, as mentioned, David Hume as well as G.W. Leibniz (1646–1716) and Christian Wolff (1679–1754). Incompatibilists divide into those who hold that determinism is false and that we have free will—the libertarians—and those who hold that determinism is true and that we lack free will—the hard determinists. In Kant's context, the pietist philosopher and theologian Christian August Crusius (1715–1775) is an influential libertarian, and for him Spinoza is the main hard determinist opponent. For several key interlocutors, such as Gottfried Ephraim Lessing (1729-1781), it is unclear whether their sympathies ultimately lie with compatibilism or with Spinoza.

3. The Problem of Free Will as Kant Sees It

The outline of the problem Kant sets out for free will, beginning in the *Critique of Pure Reaso*n (1781, 2nd ed. 1787), is this: empirically—in the realm of appearance—which he also calls nature, every event, including each of our actions, is causally determined by temporally preceding conditions:

> All the actions of the human being in appearance are determined in accord with the order of nature by his empirical character and other cooperating causes; and if we could investigate all the appearances of his power of choice down to their basis, then there would be no human action that we could not predict with certainty and recognize as necessary given its preceding conditions (A549–50/B577–78).

On Kant's proposal, empirical causal determinism does not rule out freely willed action. But he rejects any compatibilist resolution on which freely willed action is compatible with its causal history being *exhausted* by preceding natural conditions that causally determine its occurrence. In the *Critique of Practical Reason* (1788) Kant contends that such compatibilist freedom is ruled out because it cannot accommodate the control in action that freedom requires: "Since the past is no longer in my control, every action that I perform must be necessary by determining grounds *that are not within my control*, that is, I am never free at the point in time in which I act" (*KpV*, *Ak* V 94).

In this second *Critique* Kant also specifically rejects compatibilist views on which free actions are those that are determined by appropriate internal causes, as in Hume's view:

> It is a wretched subterfuge to seek to evade this by saying that the *kind* of determining grounds of his causality in accordance with natural law agrees with a *comparative* concept of freedom... [in] which the determining natural cause is internal to the acting thing... And if the freedom of our will were the latter (say, psychological and comparative but not also transcendental, i.e. absolute), then it would at bottom be nothing better than the freedom of a turnspit, which, when once it is wound up, also accomplishes its movements of itself (*KpV, Ak* V 96–7).

If an action's causal genesis were exhausted by preceding natural conditions that causally determine its occurrence, the action would not be in the subject's power in a sense sufficient for what Kant calls *practical freedom*, which

> presupposes that ... [an action's] cause in appearance was thus not so determining that there is not a causality in our power of choice such that, independently of those natural causes and even opposed to their power and influence, it might produce something determined in the temporal order in accord with empirical laws, and hence begin a series of occurrences *entirely from itself* (*ganz von selbst*) (A534/B562).

Practical freedom presupposes a power to choose independently of the natural causes that determine our actions. More exactly, practical freedom, in its negative aspect, is the power to produce an action without being causally determined by sensuous impulses (A534/B562), and in its positive aspect, which Kant also calls autonomy, it is the power to act motivated by moral principles whose source is rationality (e.g., *KpV, Ak* V 129). As Jerry Schneewind states it,

> Kantian autonomy presupposes that we are rational agents whose transcendental freedom takes us out of the domain of natural causation. It belongs to every individual, in the state of nature as well as in society (A810/B838). Through it each person has a compass that enables "common human reason" to tell what is consistent with duty and what inconsistent (Schneewind 1998: 516).

Like Hume, Kant rejects the view that freely willed actions might simply be indeterministically caused or else uncaused events in nature, as in the view of Lucretius—such events would in his view amount to "blind chance" (*KpV, Ak* V

95).[2] Instead, the sort of indeterministic causation needed is "the power of beginning a state *of itself* (*von selbst*)—the causality of which does not in turn stand under another cause determining it in time in accordance with the law of nature" (A533/B561). The agent who acts freely *of itself* begins an effect in the sensible world, and, in doing so, is not determined by preceding causes (A541/B569). Kant calls this characteristic *transcendental* freedom, and he thinks of the power of transcendental freedom as an *intelligible cause*, by contrast with an empirical cause (e.g. A537/B565).[3]

An additional element in Kant's theory of freedom is *Willkür*, "a power *to do or to refrain from doing as one pleases* (*ein Vermogen nach Belieben zu thun oder zu lassen,*" (*Metaphysics of Morals*, *Ak* VI 213), which he believes is necessary for moral obligation (to be discussed). But it is notable that for Kant, transcendental freedom need not involve having alternative possibilities for action, and it therefore does not entail having the power of *Willkür*. The following passage from the *Religion Within the Boundaries of Mere Reason* is indicative of Kant's position on this issue:

> There is no difficulty in reconciling the concept of *freedom* with the idea of God as a necessary being, for freedom does not consist in the contingency of an action (in its not being determined through any ground at all) i.e. not indeterminism ([the thesis] that God must be equally capable of doing good or evil, if his action is to be called free) but in absolute spontaneity. The latter is at risk only with predeterminism, where the determining ground of an action lies in antecedent time, so that the action is no longer in *my* power but in the hands of nature, which determines me irresistibly; since in God no temporal sequence is thinkable, this difficulty has no place (*Rel, Ak* VI 50n).

God cannot do otherwise than what is best but is yet free by virtue of divine action being produced entirely from the self *(ganz von selbst*) and thus not causally determined by conditions that precede it. For human beings, who have the power of *Willkür*, transcendental freedom may typically involve that power. But *Willkür*, as a matter of conceptual fact, is not required for transcendental freedom.

4. Avoiding Hard Determinism

But now, given Kant's empirical determinism about human action, and his indeterministic conception of free action, how can he avoid the Spinozistic result that because all actions are causally determined they are not free? Kant's solution invokes his transcendental idealism, which distinguishes between things as they appear and things as they are in themselves, independently of how they appear.[4]

[2] Here Kant writes: "If, then, one wants to attribute freedom to a being whose existence is determined in time, one cannot, so far at least, except this being from the law of natural necessity as to all events in its existence and consequently as to its actions as well; for, that would be tantamount to handing it over to blind chance" (*KpV, Ak* V 95). Lucretius's position is found in *De Rerum Natura*: "but what keeps the mind itself from having necessity within it in all actions [...] is the minute swerving of the first beginnings at no fixed place and at no fixed time" (Lucretius 1982: 2, 289-293).

[3] In the contemporary debate, agent-causal libertarians advance a position relevantly similar to Kant's, e.g. Randolph Clarke (1993) and Timothy O'Connor (2000).

[4] This component of Kant's position is set out by Allen Wood (1984: 83–89) and by Henry Allison (1990: 30–41).

While as appearances our actions are always causally determined by preceding temporal conditions, they may also originate from the self as it is in itself as their transcendentally free cause. The core features of this solution are the following. First, human agents, as appearances, have an *empirical character*. Character, Kant specifies, is a law of something's causality, where causality is the activity of a cause (A539/B567). A thing's character is the way it behaves causally. A feature of how we as appearances, as empirical selves, behave causally is that our actions are causally determined by preceding natural conditions. However, the empirical character of our actions is compatible with our actions being produced by virtue of a character of a distinct kind, an *intelligible* character (A539/B567). In producing an action an agent, as a thing in itself, as an intelligible subject, may begin a state from the self without such causal determination.

In the *Critique of Pure Reason* Kant is careful to specify that we cannot *establish* that we as intelligible subjects are transcendentally free; "we have not been trying to establish the reality of freedom", or even to show that it is *really possible* i.e., metaphysically possible, that we are transcendentally free; "we have not even tried to prove the possibility of freedom". Rather, Kant aims only to ascertain that "nature at least *does not conflict with* causality through freedom" (A558/B586). That is, to ascertain that our conception of ourselves as transcendentally free while embedded in the natural world is *not logically impossible*, i.e., that it is *logically possible*. For a conception to be logically possible is for it not to feature a contradiction (A244/B302). In Kant's view, a conception can be logically possible without its being a conception of a *real* possibility, because there are ways to preclude real possibility that do not involve establishing a contradiction. This perspective is in this respect similar to Saul Kripke's (1980) widely endorsed position according to which, for example, there is no contradiction involved in the conception that water is an element, while water's being an element, and thus not a compound, is nonetheless metaphysically impossible. Kant maintains that *we* can know (*erkennen*) real possibilities theoretically only through experience, that is, empirically, and in particular only by way of sensible intuition (A602/B630; cf. A218/B265–A226/B274), and this is akin to the contemporary view that we can know that water is a compound is metaphysically possible and that water is an element is metaphysically impossible only through experience. Accordingly, although our conception of ourselves as transcendentally free features no contradiction, this is not enough to establish that this conception is metaphysically possible.

Kant thus denies that we can know that we, as we are in ourselves, have the power of transcendental freedom. I've argued that Kant's more general denial of knowledge of things in themselves, a core element of his transcendental idealism, amounts to rejecting what I call *substance-knowledge* of them (see Pereboom 1991). Substances in the Leibnizian scheme are things in themselves (and so not appearances) whose essential features include fundamental causal powers, and these powers are conceived as intrinsic properties of these entities. Specifically, for Leibniz, all substances are monads, modeled on Descartes's immaterial souls, whose essential feature is the fundamental causal power of representation, and the monad has this power intrinsically. In denying knowledge of things in themselves, Kant is ruling out knowledge of such fundamental intrinsic causal powers of substances. Transcendental freedom would indeed be such a fundamental causal power—an intrinsic feature of human selves as they are in themselves. Accordingly, knowledge of this power is ruled out by Kant's stricture. In accord with this account, although we can form a conception of transcendental freedom, we

lack the ability to investigate the nature of fundamental causal powers of the self to establish whether transcendental freedom is a capacity we actually have, or even whether this is metaphysically possible.

But what does Kant then mean when he says he has shown that nature does not conflict with transcendental freedom (A558/B586)? What if transcendental freedom is metaphysically impossible? I suggest that the best interpretation of Kant's claim is that there is no internal inconsistency in the conception of this power that we can formulate by our reason, and that there is no inconsistency between the claim that this description is true of us as noumenal agents and our best empirical theories about the natural world. This view allows that we could never comprehend its nature as a fundamental causal power.

5. Belief in Transcendental Freedom

That this is what Kant aims to establish regarding transcendental freedom fits with his main objective—to vindicate a "belief" that we are transcendentally free. In his view, we cannot understand the fundamental causal powers of things in themselves, but we can nonetheless have legitimate beliefs about these powers. The notion of belief at issue is a "subjectively sufficient" but "objectively insufficient" conviction that is based on practical and not on theoretical—i.e. evidential—considerations (A822–3/B850–1).[5] In addition, Kant remarks that to believe in the practical sense is just to be guided practically by the content of the belief; this notion of belief "refers only to the guidance (*Leitung*) that an idea gives me, and the subjective influence on the advancement of my actions of reason that holds me fast to [the idea], even though I am not in a position to give an account of it from the speculative point of view" (A827=B855). A plausible suggestion that combines these ideas is that one has a belief in this practical sense when any reservations one might have about the truth of the proposition are set aside so that one's conviction in that proposition may guide one's thought and deliberation in acting.

Kant contends that for some such beliefs, although their evidential basis is inadequate for their justification, the practical benefits that derive from their capacity to guide our actions nevertheless justify our having them. Thus, for example, for these sorts of practical reasons one is justified in having a belief that God exists despite the absence of adequate evidential justification. At the same time, Kant indicates that in general such beliefs—at least when they are about things in themselves—should yet satisfy a theoretical requirement: they cannot be internally inconsistent or inconsistent with what we can establish theoretically. Why does he advocate this requirement? On the one hand, Kant maintains that the law of non-contradiction governs things in themselves. But still, if the practical rationality of belief is at issue, then not even inconsistency should in principle rule it out. Suppose someone with a cerebroscope credibly threatened to torture and kill me unless I were to believe some contradiction or other. Would it not then be practically rational for me to believe a contradiction (e.g., by taking a drug that would cause me to do so)? Or suppose that to believe that moral principles hold for us we would also have to believe some inconsistent proposition. Shouldn't Kant seriously consider that believing the inconsistency is then legitimate on practical grounds?

[5] For a thorough account of Kant's theory of belief, see Chignell 2007.

Perhaps Kant would seriously consider this proposal—he does not explicitly rule on the issue. But his concern is that if one is certain that a proposition is false, it won't be psychologically possible to believe it in order to secure some practical effect, and we can indeed be certain that contradictions are false. Kant discusses an example of a man lacking good moral sentiments, who, although he

> might be separated from the moral interest by the absence of all good dispositions, yet even in this case there is enough left to make him fear a divine existence and a future. For to this end nothing more is required than that he at least cannot pretend to any *certainty* that there is *no* such being and *no* future life, which would have to be proved through reason alone and thus apodeictically, since he would have to establish them to be impossible, which certainly no rational human can undertake to do (A830/B858).

If the claims to the existence of God and a future life did feature inconsistencies, then their impossibility could be established. We could then be certain that these claims were false, and if we were, it would be psychologically impossible to have the conviction in them required to secure the desired practical effect. To avoid the certainty of the falsehood of these claims, it would have to be shown that they did not feature inconsistencies.

6. Is the Claim that We Are Transcendental Free Credible?

An objection to Kant's position is that propositions can lack credibility in the sense at issue for reasons other than overt inconsistency. For instance, a proposition's being judged very highly improbable might render it at most insignificantly more credible in this sense than an inconsistent proposition. Consider the proposition that intelligent beings from another planet will conquer the earth within the next year and subject all human beings to slavery, which for me is highly improbable. This proposition features no inconsistencies, but I nevertheless I cannot believe it in the sense at issue, i.e., I cannot set aside my reservations about its truth so that my conviction in it can guide my thought and deliberation about acting.

So even if the claim that we are transcendentally free does not feature an overt contradiction, is it relevantly more credible than a proposition that does? One issue that bears on this question is whether the empirical subject that is causally determined to cause the action is to be understood as *identical* to the noumenal subject that is transcendentally free and thus not causally determined to cause that action. In the *Critique of Practical Reason* Kant explicitly asserts this identity:

> if one still wants to save [freedom], no other path remains than to ascribe the existence of a thing so far as it is determinable in time, and so too its causality in accordance with the law of *natural necessity, only to appearance, and to ascribe freedom to the same being as a thing in itself* (*KpV*, *Ak* V 95; cf. *Ak* V 97).

Many interpreters favor a *one-world* view on the relationship between appearances and things in themselves, according to which in general an appearance is numerically identical to a thing in itself or to a plurality of such things (e.g., Allison 1990). There are several key texts that lend support to the one-world view, but there are also those that might be construed to favor a *two-world* reading, according to which every appearance is numerically distinct from any thing in itself (e.g Jauernig 2012). The concern that arises on the one–world conception is that the

empirical self and the self as it is in itself would be identical while having incompatible properties. An empirical self is, in Kant's view, a complex of psychological states that can potentially be apprehended by inner sense. By his account, for any action it performs in the empirical world, the empirical self is causally determined to perform that action by preceding natural conditions. However, we are asked to believe that when the action is free, the self as it is in itself produces the action from itself without being causally determined by preceding conditions. Could the empirical self (E) and the self as it is in itself, also known as the noumenal self (N), be identical? (E) and (N) differ in their properties: (E) has the property of being causally determined by conditions beyond her control to cause the action, while (N) lacks this property by virtue of causing the action from herself. How is it at all credible that (E) and (N) should be identical while differing in this specific way?

This problem can be addressed by the two-world reading, since then E and N would not be identical. But on a two–world reading, the action, as an event in the empirical world, would be overdetermined in a peculiar way. By one strand of its causal history the empirical action has a sufficient cause in a transcendentally free self, while by another strand this same action has a sufficient cause in a natural sequence that traces back to a time before the agent came to exist. Now there would be nothing incredible about the proposal that a transcendentally free agent should make a free choice *on some particular occasion* for an action that was at the same time causally determined by a natural causal sequence. However, a more substantial proposal is required, which fares differently. It is that *all* transcendentally free choices should be for just those actions that are at the same time determined to occur by virtue of natural causal sequences, and that *none* of these choices be for alternatives to those actions. If we were selves as they are in themselves making transcendentally free choices for our actions, it seems that one would *expect*, in the long run, that these choices would be manifest in the empirical world as patterns of divergence from the deterministic natural laws. The proposal that there are no such divergences, although it involves no contradiction, would appear to run so sharply counter to what we would expect to occur as to render it incredible.[6]

Perhaps a further story can be told to make this two–world proposal credible. According to the position of Luis de Molina on divine providence, God knows, eternally, what every possible libertarian free creature would choose in every possible circumstance, and with this knowledge, God is able to direct the course of history with precision, partly in virtue of creating just those free creatures whose choices fit a preconceived divine plan.[7] On one version of this Molinist view adapted to Kant's idealism, God would reconcile noumenal transcendental freedom with phenomenal determinism by creating just those transcendentally free beings the appearances of whose free choices conform to the deterministic laws that God intends for the phenomenal world. Now Molinism is actually endorsed by many of those who have thought seriously about free will and divine providence, and this Kantian version of Molinism might not be significantly less credible. So there might indeed be some who would be able to set aside any reservations they might have about this view, with the result that the conviction that we are

[6] See Pereboom 2001: 79–85 for further development of this argument.

[7] Molina 1988. For a thorough exposition and defense of Molina's position, see Flint 1998. See also Christopher Insole's (2013) discussion, especially Chapter 9.

transcendentally free can provide guidance for their actions. But for others this Kantian version of Molinism may not be credible due to the idealistic or the theistic beliefs that it presupposes. Or else, it may not be credible because of the main objection to Molinism, that there could be no truths about what possible transcendentally free creatures would choose not-yet actual or counterfactual circumstances, since any candidate for a ground for such truths would actually only ground truths of the form: *S would probably choose A*, and *not S would in fact choose A*.[8]

7. Transcendental Freedom and Moral Obligation

Even though according to Kant it is far from true that we can establish on the basis of the evidence that we are transcendentally free, he nevertheless contends it is legitimate for us to believe that we are. The grounds for legitimacy are practical—we have reasons that derive from morality in particular for believing that we are transcendentally free.[9] Kant presents two such reasons. The first is that moral imperatives have the form of 'ought' judgments, and the truth of such judgments is incompatible with the determinism we find in nature, where each event is causally determined by preceding conditions. The second is that our judgments of moral responsibility are incompatible with determinism, or again at least with this sort of natural determinism.

On 'ought' judgments Kant claims:

> Now that this reason has causality, or that we can at least represent something of the sort in it, is clear from the *imperatives* that we propose as rules to our powers of execution in everything practical. The *ought* expresses a species of necessity and a connection with grounds which does not occur anywhere else in the whole of nature. In nature the understanding can cognize only *what exists*, or has been, or will be. It is impossible that something in it *ought to be* other than what, in all these time-relations, it in fact is; indeed the ought, if one merely has the course of nature before one's eyes, has no significance whatever (A547/B575).

But, Kant claims, "perhaps everything that *has happened* in the course of nature, and on empirical grounds inevitably had to happen, nevertheless *ought not to have happened*" (A550/B579). Or at least this is a moral claim we would assume to be true.

In the *Religion Within the Boundaries of Mere Reason*, for example, Kant explicitly endorses the "ought implies can" principle; "For from the practical point of view this idea [of a prototype of humanity pleasing to God] has complete reality within itself. For it resides in our morally-legislative reason. We *ought* to conform to it, and therefore we must *be able* to" (*Rel*, *Ak* VI: 62).[10] The following moral "ought implies can" principle is indeed attractive: if one ought to do something, then it must be the case that one can do it.[11] Accordingly, if because in general

[8] A classic source of the "grounding" objection is Adams 1979.

[9] See Markus Kohl (2023) for an extensive discussion of these issues.

[10] In addition, Kant asserts "ought implies can" at *Rel*, *Ak* VI: 45: "For, in spite of that fall, the command that we *ought* to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it"; and he defends a similar idea at *Rel*, *Ak* VI: 68: "Yet duty commands that he be good, and duty commands nothing but what we can do".

[11] For a sensitive discussion of these issues, see Sinnott-Armstrong 1984. For discussions of Kant on obligation and ability, see Stern 2004 and Kohl 2015, 2023.

one is causally determined to act as one does, one can never do otherwise, it would be false that one ever ought to do otherwise. In particular, if whenever one acts wrongly, it is never true that one ought to do otherwise, what would be the point of a system of moral 'oughts'? It would seem that if "A ought to do *x*" is true at all, it must be true not only when A does *x*, but also when A fails to do *x*.[12] If all of this is sound, then given that we have a good practical reason to preserve moral 'ought' judgments, we have a good practical reason to believe that we are transcendentally free.

But how strong is this practical reason? An initial problem is that what would be required for moral principles to be true or hold for us is not the belief that we are transcendentally free, but rather we actually are transcendentally free. By contrast, sometimes a belief itself, and not specifically the truth of the content of the belief, is what is needed to secure a practical goal. For example, in the version of the moral argument for belief in God that we find in the *Religion*, Kant's idea is that without a belief in God we could not also believe that in each person happiness will eventually be proportioned to virtue, and the suggestion is that if we did not have this belief about happiness and virtue, we would be disheartened to the degree that our motivation to moral action would suffer (*Rel, Ak* VI 6–8n; see also Adams 1979). In this case, it is the *belief* that God exists, specifically, that would prevent the hindrance to moral motivation. But, returning to the case of freedom, perhaps it is the belief that moral principles are true, rather than the truth of the moral principles, that Kant aims to secure. Given this supposition, the belief that we are transcendentally free might indeed be what is required to secure the goal.

Two worries one might raise about this proposal are: first, perhaps, in our conception of morality, there are 'ought' judgments sufficient for morality that do not presuppose an "ought implies can" principle; and second, there may be principles sufficient for morality that are not 'ought' judgments and are not undermined by an "ought implies can" principle. So first, one clear role that moral 'ought' judgments have is to guide actions. We say to people that they ought not steal, for instance, in order guide their practical reasoning so that they might refrain from stealing. Moreover, for Kant, an important function of practical judgments generally is to guide actions. Does an "ought implies can" principle need to be true for 'ought' judgments to have this action-guiding function? Not obviously. Suppose that causal determinism is true, and that hence no agent could ever done have otherwise. Nevertheless, we do not typically know in advance—before deliberation is complete or a decision has been made—which choice for action has been causally determined. Rather, it is almost always true that from the epistemic point of view of the agent at the time of deliberation more than one option for which choice she will make is possible. That is, more than one such option is possible relative to an appropriate subset of what she believes (or of what she should believe), in the sense that it is at least not ruled out by what she believes (or by what she should believe).[13] Often, when one attempts to guide an agent by means of a moral 'ought' judgment, it is the range of options for action that are in this sense epistemically possible for the agent at the time of deliberation that one addresses. Frequently, it is significantly probable that expressing a moral 'ought' judgment will causally influence action, and thus there is a good moral

[12] See Haji 2002 for a more thoroughly developed argument for this claim.
[13] See Kapitan 1986.

reason to do so—even if it turns out that because causal determinism is true the agent could not have complied with the judgment.

Against this solution one might argue that although 'ought' judgments in these action-guiding roles would retain practical value, so that it might often be practically rational to express them, they must nevertheless be false if causal determinism were true. But even Ishtiyaque who has argued at length that 'ought' judgments as instruments for deontic appraisal of actions would be false if determinism were true, agrees that this type of undermining argument does not hold for 'ought' judgments when they have an action-guiding function.[14] Haji presupposes, as C.D. Broad also contended (see Broad 1952), that 'ought' judgments have various distinct roles, and that these roles have different truth or assertability conditions. To my mind he is right to suggest that the action-guiding variety of 'ought' judgment can be retained even if determinism is true and no agent could have done otherwise, and even if this consideration undermines 'ought' judgments in other roles.

Secondly, even if moral 'ought' judgments do turn out not true or do not hold for us because our actions are causally determined, we need not also accept that no moral principles are true or hold for us. For, plausibly, moral judgments about rightness and wrongness of actions could still be true. Suppose that someone is causally determined by genetic predisposition and childhood abuse to be a violent criminal. His actions are, intuitively, still morally wrong, and it is still morally wrong for him to commit these crimes. Moreover, moral judgments such as "it is morally good for A to do *x*" and "it is morally bad for A to do *y*" still could be true. Thus, for example, even if one is causally determined to refrain from giving to charity, and even if it is therefore false that one ought to give to charity, it still might still be good to do. Embezzling funds from one's company would be a bad thing to do, even if one's act is causally determined, and hence, even if it is false that one ought not to do so. It would seem that principles regarding moral rightness and wrongness, goodness and badness, can fulfil most or even all of the roles that moral 'ought' judgments have in guiding action, in, for example, moral encouragement, admonition, and evaluation. If this is so, and since the truth of these alternative principles does not appear to require transcendental freedom, the practical reason for believing that we are transcendentally free that Kant adduces at this point would carry insufficient weight.

8. Transcendental Freedom and Blameworthiness

But even if the consciousness that there is a moral law does not all by itself support a belief in transcendental freedom, another aspect of our sense of morality does. When people perform actions contrary to the moral law, many of us typically judge them blameworthy in a sense that features basically deserved pain or harm. More specifically,

> For an agent to be *morally responsible for an action in the basic desert sense* is for the action to be attributable to her in such a way that if she was sensitive to its being morally wrong, she would deserve to be blamed or punished in a way that would be harmful to her, and if she was sensitive to its being morally exemplary, she would deserve to

[14] Haji (2002: 77) writes: "For the argument for the incompatibility of determinism and deontic morality is not in any way concerned with the action-guiding function of ought judgments".

> be praised or rewarded in a way that she would be beneficial to her. The desert at issue is basic in the sense that the agent, to be morally responsible, would deserve such blame or punishment, praise or reward, just by virtue of having performed the action with sensitivity to its moral status, and not, for example, by virtue of consequentialist or contractualist considerations (Pereboom 2021: 11–12).[15]

Basic desert moral responsibility contrasts with the sort that appeals only to non-basic desert, which invokes further goods, such as good consequences, to justify desert claims. It also contrasts with conceptions of moral responsibility that do not invoke desert at all, but are instead resolutely forward-looking, recruiting blame only to advance goods such as moral formation and restoration of relationships impaired by wrongdoing.

For many, the intuition that wrongdoers deserve to be punished concerns basic desert specifically, in contrast with its non-basic relative. In *The Metaphysics of Morals* Kant provides an example that illustrates this sense, one that features a murderer in and imagined island society that is about to dissolve itself, in which there is no further good to which punishing a wrongdoer would contribute. Kant strenuously advocates that he should be executed, just because of the crime he has committed; that is, for reasons of retributive desert alone (*Ak* VI: 331–33). To embellish the example, imagine a person on an isolated island viciously murders everyone else on the island and that he is not capable of moral reform due to ingrained hatred and rage. Thus, there are no good consequences that the punishment might aim to realize, and there is no longer a society on the island whose rules might be determined by contract. Many nonetheless have the intuition that this murderer deserves to be punished severely. The desert would be basic since the specifics of the example eliminate non-basic desert.

Accordingly, the second practical consideration Kant adduces in support of the belief that we are transcendentally free is that if we lacked this kind of freedom our judgments of blameworthiness—and moral responsibility more generally—would turn out to be false. The issue is discussed in the "malicious lie" passage in the *Critique of Pure Reason*:

> one may take a voluntary action, e.g. a malicious lie, through which a person has brought about a certain confusion in society; and one may first investigate its moving causes, through which it arose, judging on that basis how the lie and its consequences could be imputed to the person. With this first intent one goes into the sources of the person's empirical character, seeking them in a bad upbringing, bad company, also finding them in the wickedness of a natural temper insensitive to shame, partly in carelessness and thoughtlessness; in doing so one does not leave out of account the occasioning causes. In all this, one proceeds as with any investigation in the series of determining causes for a given natural effect. Now even if one believes the action to be determined by these causes, one nonetheless blames the agent, and not on account of his unhappy natural temper, not on account of the circumstances influencing him, nor even on account of the life he has led previously; for one presupposes that it can be entirely set aside how that life was constituted, and that the series of conditions that transpired might not have been, but rather that this deed could be regarded as entirely unconditioned in regard to the

[15] This is a slight revision relative to the formulation in Pereboom 2021; 11–12. For earlier formulations, see Pereboom 2001: xx, and 2014: 2.

> preceding state, as though with that act the agent had started a series of consequences entirely from himself. (A554–55/B582–83).[16]

The idea is that we have good practical reason to judge the liar blameworthy, and since blameworthiness requires transcendental freedom, we thereby have a good practical reason to believe that he is transcendentally free.

But note the epistemic situation that Kant thinks we are in. We cannot show on the basis of the evidence that we are transcendentally free, or even that transcendental freedom is causally possible, but only that a description of transcendental freedom is not internally inconsistent or inconsistent with our best theories about the empirical world. Given this epistemic situation, and assuming Kant's incompatibilism, would it be morally acceptable to judge a wrongdoer blameworthy for what he has done? Or to justify expressing one's anger towards him by the claim that he is blameworthy? Or, if he is a criminal, to deprive him of his liberty or life on the ground that he deserves such treatment just by virtue of having done wrong?

Consider, for example, the murderer in the imagined island society that is about to dissolve itself, who Kant believes basically deserves to be executed. Imagine that the murderer protests that he was determined by empirical causes to act as he did. Would the following reply be morally acceptable? "Although we have no evidence that you are transcendentally free, and although we cannot establish that such a power of agency is even metaphysically possible, yet our belief that you are free in this way involves no inconsistency, and we need to have this belief in order to justify treating people like you as blameworthy and basically deserving of punishment". This is dubious. Treating an offender as basically deserving of the harm or pain involved in blaming and punishing him requires a high standard of justification—much higher than the standard of consistency Kant endorses. If one's justification for the harm or pain depended on the claim that we are transcendentally free, and we have little or no evidence for this attribution, and the story we need to tell to reconcile transcendental freedom with our best empirical theories at best marginally credible, then that justification would appear to be inadequate.

In addition, if we relinquish basic desert, we yet have access to a model for holding moral responsible that arguably can serve the same functional role but does not presuppose transcendental freedom. Rather than confronting wrongdoing by blame suffused with retributive sentiments, we can turn to a forward-looking stance of moral protest instead. Pamela Hieronymi has proposed that blame should be understood as moral protest (see Hieronymi 2001: Chpt. 2). But while in her view the negative reactive attitudes have an essential role in blaming as protest, we can set this part of her account aside. Moral protest can be viewed as a psychological stance, a posture of mind that has certain aims or functions that are manifest as dispositions to act. It is a stance of opposition to specific immoral actions and their

[16] The "malicious lie" passage continues: "This blame is grounded on the law of reason, which regards reason as a cause that, regardless of all the empirical conditions just named, could have and ought to have determined the conduct of the person to be other than it is. And indeed one regards the causality of reason not as a mere concurrence with other causes, but as complete in itself, even if sensuous incentives were not for it but indeed entirely against it; the action is ascribed to the agent's intelligible character: now, in the moment when he lies, it is entirely his fault; hence reason, regardless of all empirical conditions of the deed, is fully free, and this deed is to be attributed entirely to its failure to act" (A555/B586).

general type, whose aims include communication of this opposition to wrongdoing, together with reasons to refrain from it. Such a model can specify that the goals of blaming are the forward-looking aims of moral formation of wrongdoers, protection against wrongdoing, and reconciliation in relationships impaired by wrongdoing.

This model can also accept the Kantian proviso that it is the agent's faculty of recognition and responsiveness to reasons that is engaged in moral protest. In that process we may request an explanation with the intent of having the agent acknowledge a motivation to act wrongly, and then, if he has in fact so acted without excuse or justification, we may intend for him to come to see that the motivation issuing in the action is best eliminated. This change is produced by way of the agent's recognition of moral reasons to make it. More generally, it is a wrongdoer's responsiveness to reasons together with the forward-looking aims that explains why he is appropriately addressed by moral protest so characterized. This model of holding morally is not tied to transcendental freedom because given its forward-looking aims, it is compatible with causal determinism.

9. Transcendental Freedom as a Fact of Reason

Kant pursues what may appear to qualify as a more ambitious route to affirming transcendental freedom in the *Critique of Practical Reason*. In this the second *Critique* Kant argues that the reality of our having freedom in this sense is established as a *fact of reason—Faktum der Vernunft* (*KpV* 5: 47), and he specifies that we have a practical cognition (*praktischen Erkenntnis*) of this fact (*KpV* 5: 103). On Andy Reath's (2024) account, it is in proper exercises of our power of reason in moral judgment that the fact of reason is given to us, and this fact is *the reality of pure practical reason and the authority of the moral law*. The proper exercises of the power of reason are instances of moral judgments that assign deliberative priority to specific moral grounds, for example truthfulness over personal advantage. Such judgments affirm the Categorical Imperative, specifically the formula of universal law—act only on that maxim that you can at the same time will as a universal law—as their formal principle. This given fact of reason yields practical cognition of the authority of the fundamental principle of morality. The reality of free will is subsequently given through the fact of reason, also as a practical cognition. The moral judgments at issue are judgments of moral obligation, and because 'ought' implies 'can', they presuppose that in cases in which one has not done what one ought, one could have done otherwise.

The account of the *Critique of Pure Reason* differs in that it invokes neither the notion that freedom is given as or via a fact of reason, nor that it is a practical cognition. The account of the second Critique might at least initially be conceived as a daring new move, aiming to show, as Karl Ameriks (2000: 189–233) puts it, *that* we are transcendentally free. But as Henry Allison (1990) points out, there are at least two reasons for restraint that derive from Kant's views more generally on any interpretation of the claim that the reality of free will is secured as or by a fact of reason. One is that Kant affirms, even in the second *Critique*, that there are no instances in which we can ascertain that we are motivated to act on the basis of the moral law, by contrast with self-interested incentives (*KpV* 5: 47). Accordingly, it is not in moral decision or action itself that the reality of practical reason and the authority of the moral law are given to us as a fact, but rather in judgments as to how the moral law enjoins us to decide and to act. In turn, our transcendental freedom is not revealed in our deciding or acting morally, but as a presupposition of the

authority of the moral law. When I recognize the authority of the moral law in cases in which I acted wrongly, I see that this recognition presupposes that I could have acted otherwise.

A second reason for restraint is that establishing the reality of practical reason and the authority of the moral law does not issue in a theoretical cognition, but rather in a practical cognition. But what exactly is a practical cognition? Kant specifies that, unlike a theoretical cognition, a practical cognition does not involve an intuition:

> Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out (*herausvernüftlen*) from antecedent data of reason, for example, from consciousness of freedom (since this is not antecedently given to us) and because it forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition, either pure or empirical (*KpV* 5: 31).[17]

As a consequence, such a practical cognition does not meet the standard of knowledge of synthetic judgments central to the critical project, i.e., that it is based on intuition.

Kant makes apparently stronger and weaker claims about the epistemic status of the fact of reason. For an apparently stronger claim, he writes in *Critique of Practical Reason* that in the first *Critique* "theoretical reason was forced to assume at least the possibility of freedom to fill a need of its own", but now states that "the moral law proves its reality […] by the concept of reason determining the will immediately" (*KpV* 5: 48). Yet when Kant comes to stating precisely what he means, he is circumspect. The concept of causality as it applies to freedom has reference only to the practical, and "beyond this they lay no claim to the cognition of [intelligences]". Moreover,

> as for whatever other properties belonging to the supersensible way of representing those things may be brought forward in connection with these categories, these are without exception to be counted not as knowledge [*Wissen*] but only as warrant [*Befügnis*] (for practical purposes, however, a necessity) to admit [*anzunehmen*] and presuppose [*vorauszusetzen*] them, even where supersensible beings (such as God) are assumed by analogy, that is by purely rational relation of which we make a practical use with respect to what is sensible; and so, by this application to the supersensible but only for practical purposes pure theoretical reason is not given the least encouragement to fly into the transcendent (*KpV* 5: 57).

Here is an interpretation that I find attractive. When we take the practical point of view in making moral judgments, the content and authority of the moral law are, from the perspective of this moral consciousness, presented to us as a fact. Implicit in this fact, and derivable from it, is the judgment that we are transcendentally free. But this does not provide theoretical knowledge that we are transcendentally free—this would require an intuition that is unavailable to us. From the theoretical point of view the fact of freedom given to us practically does not amount to knowledge (*Wissen*), but rather to a practically warranted presupposition.

[17] See also (*KpV* 5: 55): "The objective reality of a pure will or, what is the same thing, of a pure practical reason is given a priori in the moral law, as it were by a fact—for so we may call determination of the will that is unavoidable even though it does not rest on any empirical principles".

On this interpretation, the accounts of the two *Critiques* are consistent. Both agree that we have no intuition and no theoretical knowledge of our transcendental freedom. But the *Critique of Practical Reason* adds that from the practical perspective, when we are making moral judgments, the moral law and its authority is presented to us as a fact, as is, derivatively, our transcendental freedom. Furthermore, we are practically justified in taking the authority of the moral law and human transcendental freedom as presuppositions in moral deliberation.

10. Final Words

Kant's theory of transcendental freedom stands out as a signature attempt to reconcile our views of ourselves as free and responsible agents with an increasingly sophisticated scientific picture of the world, and it has been an important philosophical inspiration and source for many who have this ambition. As a rule, determinists about nature are either hard determinists or compatibilists about free will. But these positions arguably relinquish widespread intuitions about the capacities for action we have, or about what is required for moral responsibility. Kant's theory is especially ambitious in that it aims to preserve these intuitions by developing a view of free will on which we, as substances, have the power to cause actions without being causally determined to cause them. But he maintains that this cannot be established or even shown to metaphysically possible on the basis of evidence we available to us. Instead, he claims only that we can show that our conception of our having free will in this sense involves no inconsistency, and that the legitimacy of a belief that we have this kind of freedom must rely on practical reasons.

To my mind, there are several important respects in which Kant's treatment of free will and action is highly plausible. Moral responsibility must indeed be grounded in transcendental freedom, but whether we have it cannot be established on evidence available to us. Yet Kant provides a consistent conception according to which we have free will of this sort. In addition, the practical reasons that Kant adduces for belief in transcendental freedom—that it is required for moral obligation and for blameworthiness—carry significant weight. However, Kant's justification for believing that we are transcendentally free based on these reasons is subject to challenge. But there may be more to these reasons than I have seen, and there are other reasons for believing in transcendental freedom that may also need to be considered. Moreover, the credibility and commitments of alternative positions on free will must be weighed against Kant's. In consequence, the general contours of Kant's position on free will remain a significant and attractive option.

References

Kant's works

Citations of '*Ak*' refer to Immanuel Kant, *Kant's gesammelte Schriften*, edited by the *Königliche Akademie der Wissenschaften* and its successors (Berlin: George Reimer (subsequently W. de Gruyter), 1902–).

A/B *Critique of pure reason.* English quotations are from the translation by Paul Guyer and Allen Wood, *The Cambridge edition of the works of Immanuel Kant: the Critique of pure reason* (Cambridge: Cambridge University Press, 1997). Any alterations are accompanied by the German equivalent.

KpV Critique of practical reason. English quotations are from the translation by Mary Gregor, *The Cambridge edition of the works of Immanuel Kant: practical philosophy* (Cambridge: Cambridge University Press, 1996), unless otherwise indicated.

KU Critique of the power of judgment. English quotations are from the translation by Paul Guyer and Eric Matthews, *The Cambridge edition of the works of Immanuel Kant: Critique of the power of judgment* (Cambridge: Cambridge University Press, 2000).

Rel Religion within the boundaries of mere reason. English quotations are from the translation by Allen Wood and George di Giovanni, *The Cambridge edition of the works of Immanuel Kant: religion and rational theology* (Cambridge: Cambridge University Press, 1996).

VpR Lectures on philosophical theology. English quotations are from the translation by Allen Wood and Gertrude Clark (Ithaca: Cornell University Press, 1978).

Adams, R.M., 1977. Middle knowledge and the problem of evil. *American philosophical quarterly*, 14, 109–117.

Adams, R., 1979. Moral arguments for theistic belief. *In*: C.F. Delaney, ed., *Rationality and religious belief*. Notre Dame: Notre Dame University Press, 116–140; repr. in Adams, R., 1987. *The virtue of faith*. Oxford: Oxford University Press, 144–163.

Allison, H., 1990. *Kant's theory of freedom*. Cambridge: Cambridge University Press.

Ameriks, K., 2000. *Kant's theory of mind, new edition*. Oxford: Oxford University Press.

Broad, C.D., 1952. Determinism, indeterminism, libertarianism. *In*: Broad, C.D., *Ethics and the history of philosophy*. London: Routledge and Kegan Paul, 195–217.

Chignell, A., 2007. Belief in Kant. *The philosophical review*, 116 (3), 323–360.

Clarke, R., 1993. Toward a credible agent-causal account of free will. *Noûs*, 27, 191–203.

Flint, T.P. 1998. *Divine Providence: the Molinist account*. Ithaca: Cornell University.

Haji, I., 1998. *Moral appraisability*. New York: Oxford University Press.

Haji, I., 2002. *Deontic morality and control*. Cambridge: Cambridge University Press.

Hieronymi, P., 2001. Articulating an uncompromising forgiveness. *Philosophy and phenomenological research*, 62, 529–555.

Hume, D., 1978. *A treatise of human nature*. Oxford: Oxford University Press.

Hume, D., 2000. *An enquiry concerning human understanding*. Oxford: Oxford University Press.

Insole, C.J., 2013. Kant and the Creation of Freedom. Oxford: Oxford University Press.

Kapitan, T., 1986. Deliberation and the presumption of open alternatives. *The philosophical quarterly*, 36, 230–251.

Kohl, M., 2015. Kant and 'ought implies can'. *The philosophical quarterly*, 65, 690–710.

Kohl, M., 2023. *Kant on Freedom and Rational Agency*. Oxford: Oxford University Press.

Kripke, S., 1980. *Naming and necessity*. Cambridge, MA: Harvard University Press.

Jauernig, Anja. 2021. *The World According to Kant*. Oxford: Oxford University Press.

Lucretius, 1982. *De rerum natura.* Translation by W.H.D. Rouse. Loeb Classical Library, Cambridge: Harvard University Press.

Molina, L. de., 1988 [1595], *Liberi arbitrii cum gratiae donis, divina praescientia, providentia, praedestinatione et reprobatione*. Translation (of Part IV) by A.J., Freddoso. *On divine foreknowledge: part IV of the Concordia*. Ithaca: Cornell University Press.

O'Connor, T., 2000. *Persons and causes*. New York: Oxford University Press.

Pereboom, D., 1991. Is Kant's transcendental philosophy inconsistent? *History of philosophy quarterly*, 8, 357–372.

Pereboom, D., 2001. *Living without free will*. Cambridge: Cambridge University Press.

Pereboom, D., 2006. Kant on transcendental freedom. *Philosophy and phenomenological research*, 73 (3), 537–567.

Pereboom, D., 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

Pereboom, D., 2021 *Wrongdoing and the moral emotions*. Oxford: Oxford University Press.

Reath, A., 2024. Rational powers and the "fact of reason". *In*: W. Gobsch and T. Land, eds., *The Aristotelian Kant*, Cambridge: Cambridge University Press, 180–206.

Schneewind, J.B., 1998. *The invention of autonomy*. Cambridge: Cambridge University Press.

Sinnott-Armstrong, W., 1984. 'Ought' conversationally implies 'can'. *The Philosophical review*, XCIII, 249–261.

Spinoza, B., 1985. *The collected works of Spinoza*. E. Curley, ed. and tr., Vol. 1. Princeton: Princeton University Press.

Stern, R. Does 'ought' imply 'can'? And did Kant think it does? *Utilitas*, 16 (1), 42–61.

Watkins, E., 2005. *Kant and the metaphysics of causality*. Cambridge: Cambridge University Press.

Wood, A.W., 1984. Kant's compatibilism. *In*: A.W. Wood, ed., *Self and nature in Kant's philosophy*. Ithaca: Cornell University Press, 73–101.

Hegel on Free Will

Thomas Meyer

Humboldt-Universität zu Berlin

Abstract

In this essay I present Hegel's philosophy of free will. Although free will plays a crucial role in Hegel's practical philosophy, freedom is also part of his philosophy of mind, his philosophy of nature, and his *Science of Logic*. After examining the philosophical motivations that led Hegel to create his system of philosophy, I will outline the basic concept of free will presented in the introduction to his *Elements of the Philosophy of Right*. This concept, however, still allows for free will skepticism, which motivates me to reconstruct the metaphysical underpinnings of Hegel's understanding of free will in his *Science of Logic*, which I will call his metaphysics of free will. In doing so, I will show that Hegel attempted to deal with the problem of the (in)compatibility of free will and the causal determination of the world. Finally, I present what I call Hegelian libertarianism, a form of an incompatibilist and indeterminist position on free will.

Keywords: Arbitrary will, Causality, Content, Self-causation, Metaphysics.

1. Introduction

Free will undoubtedly plays a significant role in Hegel's practical philosophy as a whole.[1] The intricate connection between free will and practical normativity is widely accepted such that free will is often taken to be a necessary condition for all practical normativity. It seems to make no sense to claim that we should or should not perform certain actions if we are (in principle) incapable of acting or refraining from acting otherwise. What's more interesting is the fact that the concept of freedom also plays a key role in parts of Hegel's philosophical system that are less frequently associated with the concept of free will—at least in his *Science of Logic* and the whole *Philosophy of Mind*, that is, in parts of his system that do not deal with questions of practical philosophy.

[1] Building on the basic normative concept of "right", Hegel defines "right" as "the *existence* of the *free will*" (Hegel 1991: 58 [§29]). For specialists and those who may read Hegel in other editions or languages, I give the section numbers in brackets after the page numbers in my citations to the modern translations of the works referenced. In these additional citation details, R = remark, A = addition.

Argumenta 11, 1 (2025): 111–128
ISSN 2465-2334

DOI 10.14275/2465-2334/20240.mey
First published: 12 March 2024

And yet, despite the fact that Hegel's concept of free will is central to his philosophy, it is very difficult to understand. And for this reason, or so I think, it is not well known outside of Hegel scholarship.[2] My goal in this essay, then, is to make Hegel's position explicit and accessible to those unfamiliar with his philosophy. Through (2) an examination of Hegel's philosophy, his method, and some metaphilosophical principles, (3) an analysis of key passages of the *Elements of the Philosophy of Right*, and (4) a discussion of what I call Hegel's metaphysics of free will, I explain what he has to say about the basic phenomenon of free will as the ability to bring about changes in the world at will.[3] I conclude (5) with a brief characterization of a Hegelian libertarianism and hint at the fact that Hegel's core understanding of free will is enriched by a social philosophy.

2. Overview of Hegel's Philosophy

At least two facets of Hegel's philosophy make it very difficult to reconstruct his position on any classical philosophical problem, such as free will:

(1) Hegel's philosophy consists of a whole system, so that each topic has its very specific place.[4]

(2) The way in which traditional problems and positions are incorporated into this system transforms them in such a way that they lose their traditional form.[5]

Hegel not only criticized certain positions, but also the way in which a philosophical problem is formulated in the first place. It is this second facet of his philosophy that makes it particularly difficult to understand Hegel's account of free will and its relation to determinism. In order to overcome this difficulty and to reconstruct Hegel's theory of free will, one must first understand Hegel's system of philosophy and some of its central (metaphilosophical) tenets.

Hegel's account of free will consists of three sets of doctrines: (1) his metaphysics of free will; (2) his concept of free will as self-determination (in its forms of natural will, arbitrary will, and free will); (3) his theory of 'social' freedom, which consists of institutions that allow degrees of freedom to be realized. The basic structure of Hegel's entire philosophical system consists of three parts.[6] The

[2] For example, anthologies and historical surveys of positions on free will include Aristotle, Plato, and Kant, but not Hegel. Similarly, those on topics such as causality.

[3] This directly contradicts the reading of Pippin in Pippin 1999. I will discuss this disagreement in section 4 of this paper. For a critique of Pippin's reading, see Rorty 1999.

[4] With regard to free will, Hegel explicitly asserts: "The deduction *that* the will is free and of *what* the will and freedom are—as already remarked in §2 above—is possible only within the context of the whole [of philosophy]" (Hegel 1991: 36-37 [§4R]).

[5] The two facets are independent of each other: it is possible to have a system without criticizing the presuppositions of certain formulations of a problem (see, for example, the system in Lewis 1973: 88). And one can criticize the formulation of a particular problem without arguing for a whole system of philosophy (see Ryle 1949/2009: 1-13). For an interesting similarity between Ryle's and Hegel's solutions to the mind-body problem, see Wolff 1992.

[6] Hegel's system as a whole can be found in the *Encyclopedia of the Philosophical Sciences in Basic Outline* (Hegel 2010a; Hegel 1970/2004; Hegel 2007).

first part, the *Science of Logic* (Hegel 2010b), contains Hegel's metaphysics.[7] The second part contains his *Philosophy of Nature* (Hegel 1970/2004), and the third part his *Philosophy of Spirit* (Hegel 2007). Not only does his system consist of these three parts, but Hegel claims a continuous development of all phenomena from beginning to end. Moreover, this continuity of concept development is said to be necessary. In addition, each concept can be reused to analyze and explain the next concepts. This is why the first facet—its systematicity—makes Hegel's philosophy so difficult to understand. One element cannot be understood without the others. Look at the passages on, say, the concept of free will, and you will be led to other passages involving central categories of the system. Those passages will lead you to other passages, and so on.

A first step in understanding Hegel's view of free will is to look at some rare and scattered passages in his (systematic) work[8] where he comments on the traditional problem of free will. In these passages one can find clues as to where to look in the system to learn about Hegel's own view.

In one such remark in the *Philosophy of Right*, for example, Hegel criticizes the concept of arbitrary will, which I will discuss in more detail below. The problem with understanding free will as arbitrary will (allegedly presupposed, according to Hegel, by the traditional understanding of the problem of free will) is that it is dependent on externally given contents of the will. Even if I can choose whether or not to eat, given that I am hungry, I cannot choose to be hungry as such. The fact that I am hungry is something that is caused by a physiological state of my body. One criterion that a theory of free will must meet, then, is to explain how a willing being can determine the content of its own will. The content of a truly free willing being's will must be "its own self-determining activity as such" (Hegel 1991: 49 [§15R]). Hegel asserts that "arbitrariness may indeed be called a delusion if it is supposed to be equivalent to freedom" (Hegel 1991: 49 [§15R]).

Another passage in the *Philosophy of Right* has to do with Hegel's criticism of proving freedom of the will by reactive attitudes[9]. He writes

> [I]n the manner of the older empirical psychology, the so-called *proof* of the will's freedom was derived from the various feelings and phenomena [*Empfindungen und Erscheinungen*] of ordinary consciousness, such as remorse, guilt, and the like, which could allegedly be *explained* only in terms of a *free* will" (Hegel 1992: 36 [§4R]).

Hereafter Hegel makes clear that he takes this to be the wrong way to analyze and justify freedom of the will. These passages therefore suggest a list of criteria that,

[7] I use the term 'metaphysics' in a very broad and inclusive sense, as the study of the most fundamental categories or phenomena of reality, such as identity, unity, substance, or causality. For a more recent discussion of the status of Hegel's *Science of Logic* as metaphysics, see the contributions in De Laurentiis 2016 and Illetterati and Miolli 2022 and their further bibliographic references.

[8] In addition to the *Encyclopedia* mentioned above, the systematic oeuvre also includes the *Science of Logic* (Hegel 2010b) in three books as an extended version of the first part of the system, and the *Elements of the Philosophy of Right* (Hegel 1991) as an extended version of the second part of the *Philosophy of Spirit*, which Hegel calls "Objective Spirit" (Hegel 2007).

[9] The term 'reactive attitude' goes back to Peter Frederick Strawson and his highly influential paper *Freedom and Resentment* (1962/2008).

in Hegel's view, a philosophical account of free will must meet. The first two criteria read as follows:

C1: A philosophical account of free will must show that and how a will can determine the content of its willing itself, instead of merely choosing between externally given contents of willing.

C2: A philosophical account of free will must not use phenomena that depend on the reality of free will (such as reactive attitudes) as an argument for the reality of free will.

As I will show later in this article, there are two more criteria to be mentioned, which are the focus of this essay:

C3: A philosophical account of free will must allow for alternative possibilities as a necessary, though not sufficient, condition for free will.

C4: A philosophical account of free will must explain metaphysically how free will fits into a world of external, causal determination.

I will argue C3 can be deduced out of a passage whereas the argument for C4 is rather indirect in that Hegel is committed to it just by the way his whole system is constructed. I will start now by interpreting decisive passages of his *Philosophy of Right*.

3. Hegel's Concept of the Free Will

The most comprehensive and explicit account of Hegel's understanding of free will can be found in the introduction to his *Elements of the Philosophy of Right* (Hegel 1991: 25-64 [§§1-33]). According to my reading of the introduction,[10] this presentation is structured as follows. In §§5-9, Hegel presents the basic concept of the will.[11] He then introduces three forms of will, the natural will (§§10-13), the arbitrary will (§§14-20), and finally the free will (§§21-24).

Two things are important to note about this section of the Introduction. *First*, Hegel distinguishes between the concept and the phenomenon of the will. *Second*, Hegel distinguishes between the will as it appears to us in everyday life and the philosophy of the will, which uses a special philosophical terminology to analyze the true structure of the phenomenon. Hegel himself mentions that "it is possible to form an idea [*Vorstellen*] of them by consulting the self-consciousness of any individual. In the first place, anyone can discover in himself" (Hegel 1991: 37 [§4R]) certain abilities that are forms of a will.

To begin with what everyone can "discover" in themselves, one must look at Hegel's phenomenological description of willing.[12] Hegel distinguishes two different capacities that every willing being can experience in its everyday actualizations of will. There is the "ability to abstract from anything whatsoever, and likewise to determine himself, to posit any content in himself by his own agency" (Hegel 1991: 37 [§4R]). Thus:

Ability$_1$: To be able to abstract from any possible propositional content of willing.

[10] I present a more explicit analysis of the introduction in Meyer 2020, Ch. 1.

[11] Contrary to the standard reading of the Introduction (like in Quante 1997 and Vieweg 2020), I also count §§8-9 as part of the basic concept of the will, because only there does Hegel present the difference between form and content of the will.

[12] For a phenomenology of free will, see Nahmias et al. 2013.

Ability$_2$: To be able to determine oneself by deciding for a propositional content of willing.

As I discuss below, both capacities are interdependent for their exercise. We can neither decide what we want to do without abstracting from the given possible contents of willing and without reflecting on these contents, nor can we exercise our ability to will if we would never finally decide for something specific. For, to abstract means to *not* do which means to *not* decide. What is Hegel's philosophical analysis of these aspects of willing?

3.1 Hegel's Abstract Concept of the Will

In what Hegel calls the "*formal* [*formale*] will" (Hegel 1991: 43 [§8]), he distinguishes different aspects of each act of will. *First*, there is the distinction between subjectivity and objectivity. As a presupposition of every intention, there is a subjective, inner will that is opposed to an objective world, so that what is willed is not yet objectively the case. We, as willing beings, are self-conscious subjects who find ourselves in an objective world outside of ourselves. Our inner decision for a certain to-be-realized content of our will refers to this external world, for it is this objective world in which the content of the will is to be realized. But it is not realized as long as it is only part of the subjective, inner will. Therefore, and this is a *second* aspect, willing consists in the process of realizing this content. Willing means at least trying to realize one's intention. *Finally*, especially if one succeeds, the mere subjective content of willing has become an objective content of willing, a state of affairs. This description is obviously formal: it is absolutely neutral with regard to the content of one's will. Will, then, is constituted by the opposition between the subjective and the objective, and by the process of realizing the subjective in the objective.

The term "content of the will" then is introduced as such: "In so far as the will's determinations are *its own*—that is, its *internally* reflected particularization in general—they are its *content*" (Hegel 1991: 43 [§9]). Hegel then introduces the notion of purpose to refer to the content of the will. He distinguishes between the subjective purpose "as represented in the act of willing" and the objective purpose, "its end as actualized and accomplished through the mediation of its activity as it translates the subjective into objectivity" (Hegel 1991: 43 [§9]). I will therefore use the term "end" to refer to what is the content of one's will. The process of attempting to realize the end can thus be called an "end-realization-attempt" (Gethmann 1998, 2016, my translation).[13]

In order to make an end one's own, one must be able to abstract from different ends. This ability$_1$ in turn depends for its possibility on what Hegel calls "universality" (Hegel 1991: 37 [§37]). "The will contains (α) the element of *pure indeterminacy* or of the 'I''s pure reflection into itself, in which every limitation, every content, whether present immediately through nature, through needs, desires, and drives, or given and determined in some other way, is dissolved; this is the limitless infinity of *absolute abstraction* or *universality*, the pure thinking of oneself" (Hegel 1991: 37 [§5]).

All of these admittedly dubious characterizations are based on the idea that will is a form of mind. As volitional beings, we are spiritual beings. This sense of the spiritual can be explained by the ability to think—which in turn requires the

[13] Gethmann uses this term to define 'action'.

ability to know that one is thinking (i.e., to have self-consciousness). This in turn includes the ability to abstract or refrain from mentally represented possible ends—thus ability$_1$. Pure thinking about oneself is an alternative description of self-consciousness. Insofar as will consists among other things of this ability$_1$, which is a form of self-consciousness, a necessary condition holds for the will:

For all *x*: *x* has a will only if *x* has self-consciousness.

Of course, our will would not be a will if we were not able to decide on a certain end, to choose among many possible ends the one we want to realize. This aspect is the particularity of the will:

For all *x*: *x* has a will only if *x* determines itself, particularizes itself.

Through this moment of self-determination, "'*I*' steps into existence" (Hegel 1991: 39 [§6]). Formulated in first-person terms, this means that *I* only come into existence as this concrete willing subject, which *I* am, when *I* determine myself to a content of my will. As the addition to §6 states: "I do not merely will—I will *something*" (Hegel 1991: 40 [§6A]).[14] It is precisely this ability that most participants in contemporary debates about free will have in mind when they think about the ability to choose.

Up to this point, Hegel's description seems to reflect the way we perceive ourselves as willing beings. Only in §7 does it become clear that universality and particularity for Hegel are merely abstractions and thus one-sided moments of self-consciousness, whose unity is the will, the unity of

> *particularity* reflected *into itself* and thereby restored to *universality*. It is *individuality* [*Einzelheit*], the *self-determination* of the 'I', in that it posits itself as the negative of itself, that is, as *determinate* and *limited*, and at the same time remains with itself [*bei sich*], that is, in its *identity with itself* and universality; and in this determination, it joins together with itself alone (Hegel 1991: 41 [§7]).

This passage is difficult to understand. In a sense, the particularity and universality of the will contradict each other, since the former is about the will being undetermined, while the latter is about the will being determined. How an act of will can be understood as still maintaining the first aspect of universality is described by Hegel through his core understanding of the freedom of the will:

> "I" determines itself in so far as it is the self-reference of negativity. As this *reference to itself*, it is likewise indifferent to this determinacy; it knows the latter as its own and as *ideal*, as a mere *possibility* by which it is not restricted but in which it finds itself merely because it posits itself in it.—This is the *freedom* of the will, which constitutes the concept or substantiality of the will, its gravity, just as gravity constitutes the substantiality of a body (Hegel 1991: 41 [§7]).

For example, suppose a person *A* is thinking about certain given possible actions in an upcoming situation. After a while, *A* decides on a certain way of acting (Φ) and starts to act. Even though *A* becomes determined, that is, willing and doing Φ, this determination is not a constraint on *A*'s universality, because (1) it was *A* herself who decided for Φ, and (2) because *A* did not necessarily decide for Φ.

[14] The additions (*Zusätze*) were added to the main text after Hegel's death and consist of lecture notes from Hegel's students.

Deciding for Φ was an alternative, and A could have decided and acted differently.

Hegel further explains his understanding of freedom of the will through the concept of self-determination. In other words, when we determine our will and decide to do something, we limit ourselves. For not only do we decide that we want to do Φ, but at the same time we decide against doing all other alternative actions Φ' that are incompatible with doing Φ. In doing so, however, we set a limit for ourselves. Because we, as self-conscious beings, can abstract from everything and thus enter into the thinking-contemplative distance from everything, we must also preserve this characteristic of our universality in self-determination. We do this by staying with ourselves in self-determination. I am the one who decides. And this I, which decides, remains at the same time according to its concept, even if it determines a certain content. The reference to ends, to determinations of the will, is a possible reference of the kind that one can know in principle that one does not have to choose a certain end. It is only possible to decide for it. Therefore, the will is indifferent to its ends.

This claim about the self-determination of our will implies that Hegel's understanding of freedom of the will is committed to alternative possibilities. Suppose I am considering whether to Φ or ψ, where Φing and ψing are strictly opposed. Having decided for ψ and thus against Φing, it still holds that I could have done differently, that I could have decided for Φ and thus against ψing.[15] As one discussant in a dialogue on free will puts it, "being able to act differently is the same as being confronted with alternatives" (Williams 1980: 36).

3.2 Three Forms of the Will: Natural, Arbitrary, Free

On the first level, the content of the will is "*immediate*" (Hegel 1991: 44 [§10]). Hegel calls this first form of will "*natural* will". In this form of will it is "the *drives*, *desires*, *and inclinations* by which the will finds itself naturally determined" (Hegel 1991: 45 [§11]). On the level of the natural will, these determinations appear to be immediately given. Insofar as I have a natural will, there is still a lack of rationality. Form and content are different: the form is that everything is for me, that it is I who wills. The content, however, is not yet me, but immediately given, natural content. So the form of the will is that the content is for me, my content. Such a finite will is therefore rational in itself, but "does not yet have the form of rationality" (Hegel 1991: 45 [§11]). Here it is already clear that Hegel himself gives space to the drives and thus to the natural will within the framework of his philosophy of right. There is something objective about the drives that is "without the form of irrationality" (Hegel 1991: 45 [§11R]). Especially with regard to the second form of will, the arbitrary will, it may help to emphasize that the natural will does not yet choose between different options. What makes it still a form of will is the actual decision for a certain content. For those familiar with the so-called Frankfurt-style cases in the analytic philosophy of free will, natural will is

[15] Of course, the mere claim of alternative possibilities does not in itself prove that Hegel is right, but it does show that his concept of freedom of the will implies alternative possibilities and thus the ability to do otherwise, therefore criterion C3. I discuss the metaphysical underpinnings of this claim and the direction of the argument below.

precisely what saves freedom from cases where there is no alternative. It is the mere fact of the decision that makes it a form of will.[16]

It is not easy to understand what Hegel has in mind when he speaks of the natural will. It seems to be a borderline case, such that small children and perhaps some animals might have a natural will in this sense. He calls it "natural" because the content is naturally given, and yet it is a case of will because there is the element of decision. At the same time, the decision must not be confused with a choice between at least two different options. Natural will does not decide between options, there is no reflection on given contents. And yet the actual decision for a content and the action on it is a necessary part of the will. Although it may still be difficult to understand, I would suggest that the exercise of natural will is different from, say, a mere disposition or reflex, a naturally caused movement, for example, as is the difference between what, say, a dog did and where something happened to the dog.[17]

A limitation of the natural will arises, a function of the formal character of the natural will. Within the natural will, form and content fall apart. Therefore, the natural will participates in freedom only insofar as it is a deciding will. The fact that the natural will decides at all makes it partly free, but it is still not free insofar as its content is not determined by itself. Not only are the possible ends naturally given, but the natural will does not reflect upon and compare these different possibilities. Rather, the main exercise of the will consists in the affirmative decision for a certain content. Insofar as free agency is understood as the ability to decide for an end and to act on this decision (see Dworkin 1970: 381), natural will can also be called the ability to *act freely*.

As natural, the will fulfills only the first conceptual moment of being "the self-reflecting *infinite* '*I*' which is with itself [*bei sich selbst*] (see §5)" (Hegel 1991: 47 [§14]). In terms of content, the natural will is not with itself, but in natural drives, desires, and needs as something other than itself. Due to the structure of self-consciousness, however, finite will is "*above* its content" (Hegel 1991: 47 [§14]). In one respect it is bound to this content "as to the determinations of its nature and of its external actuality", namely as "only formally infinite" will; unbound, however—and insofar as it stands just above the content—it is as "indeterminate" (Hegel 1991: 47 [§14]) will. It follows from the first conceptual feature that the content of the will, which the will makes its own, is only one possible content among others:

> [S]ince it is indeterminate, it is not restricted to this or that content in particular. To this extent, this content is only a possible one for the reflection of the '*I*' into itself; it may or may not be mine; and 'I' is the *possibility* of determining myself to this or to something else, of *choosing* between these determinations which the 'I' must in this respect regard as external (Hegel 1991: 47-48 [§14]).

The immediately given contents of the will are then possible contents of the will insofar as it depends on me, on my choice, whether I set them as mine or not.

[16] Pippin explicitly rejects this element of the will as part of a free will in Pippin 1999: 194. I will discuss Pippin's view in the section on Hegel's metaphysics of freedom.

[17] In a sense, Wittgenstein's famous distinction between "I raise my arm" and "my arm rises" (Wittgenstein [1953] 2009, §621) can also be applied to small children and animals.

Furthermore, I can choose among the contents: "The freedom of the will, according to this determination, is *arbitrariness*, in which the following two factors are contained: free reflection, which abstracts from everything, and dependence on an inwardly or externally given content and material" (Hegel 1991: 48 [§15]). If one further defines natural will in such a way that it is itself that chooses a content, then the freedom of will consists in arbitrariness (hence *arbitrary* will). This leads to the problem of happiness for free will-indeterminism. "Since this content, which is necessary *in itself* as an end, is at the same time determined as a possible content in opposition to free reflection, it follows that arbitrariness is *contingency* in the shape of will" (Hegel 1991: 48 [§15]). Insofar as arbitrary will is understood as freedom of choice, it is contingency:

(P1) The natural content of the will is in itself a necessary content.
(P2) As an object of reflection, this content is only possible or not necessary content.
(P3) The alternation of necessity and possibility is contingency.
(C) Therefore arbitrariness is contingency, since it alternates between necessary and non-necessary content.

Obviously, the first two premises are contradictory. How could we interpret them in a consistent way? That the natural content of the will is necessary could mean that I feel hunger, pain, or tiredness because of the necessity of a natural law. It is not necessary or possible in the sense that it is up to me how to respond to this natural content of the will. Interestingly, for Hegel, contingency is not a possibility that is not necessary. Rather, contingency consists in a certain combination of naturally necessarily given contents and the possibility of reacting to these contents in one way or another. What is contingent in the classical sense is that what the arbitrary will wills is something actual even though it is not necessary. So I interpret the conclusion as follows: There is no principle that decides for a certain given content.

Hegel also identifies the shortcomings of this second understanding of the free will. It is true that reflection, as the knowledge of one's own freedom of choice, brings with it a certainty that enriches the natural will. But the problem is that free will itself is not the object of will. Instead, the object of the will is an externally given content. Thus the object is not yet an infinite one, as would be the case if the will had itself as its content. Therefore it follows: "Instead of being the will in its truth, arbitrariness is rather the will as *contradiction*" (Hegel 1991: 48 [§15R]). The argument for the thesis of the contradictory nature of arbitrary will is as follows:

(P1) Free will is absolute self-determination.
(P2) Free will as arbitrariness is self-determination according to a given content and thus not absolute self-determination.
(C) Free will as arbitrariness is absolute and not absolute self-determination.

Because of the conceptual moment of universality, in order to be able to refrain from all content, it is also possible to refrain from a chosen content. Thus it might seem that the freedom of the will corresponds to the concept of infinity. However, it "does not escape from finitude" (Hegel 1991: 50 [§16]). The mere renunciation of given finite contents testifies to independence only to a certain degree and thus allows autonomy only in a limited sense. However, this step of the abstract negative is itself afflicted with the negatum, since "every such content is different from

the form [of the will] and therefore finite; and the opposite of determinacy—namely indeterminacy, indecision, or abstraction—is only the other, equally one-sided moment" (Hegel 1991: 50 [§16]). As long as I can preserve my infinity only by renouncing the given contents of the will that are available for choice, I am not really free, for the ability to make a particularizing decision would be lacking. Without the ability to make a concrete choice, I remain only abstractly free. I call this kind of freedom *particular autonomy*. We are autonomous in the exercise of our arbitrary will insofar as we recognize our particular, contingent nature.

Hegel's discussion of possible conflicts between given contents leads us to what real or, as I call it, *universal autonomy* would look like. There are all sorts of different impulses, needs, and desires that drive us to do or not to do various things. Let us call these natural (practical) reasons. In many cases, the following conflict of incompatibility of natural reasons applies:

> Two NATURAL REASONS (drives, needs, etc.) *x* and *y* are INCOMPATIBLE with each other *iff* it is not possible to satisfy both *x* and *y* at the same time.[18]

For example, if "the satisfaction of the one demands that the satisfaction of the other be subordinated or sacrificed" (Hegel 1991: 50 [§17]), then these are two incompatible natural reasons. This observation may not seem new, since everyone knows this kind of conflict. On the level of freedom of choice, however, the question arises as to how such cases are and can be dealt with. According to Hegel, dealing with conflicts of natural reasons consists in mere decisionism. It is an arbitrary decision for which there is no additional decision principle, except to decide precisely. Although Hegel names one possible decision principle, the maximization of satisfaction, the existence of such a principle does not change the fact that the decision is decisionistic, since in the end the choice of one rather than the other decision principle is—again—left to mere decision.[19] It is true that the totality of the given drives and needs can be brought into an order, but even so, arbitrariness is not an adequate determination of free will, since it only fulfills "*formal universality*" (Hegel 1991: 52 [§20]).

What is interesting, however, is that here one might get a first idea of what real free will consists of. The difference between arbitrariness and free will would be the difference between choosing a certain flavor of ice cream and choosing a certain principle of decision. The fact that I choose a certain principle of decision is not something that came up naturally, like my desire to eat ice cream. It is something that I made up myself. In this sense, free will proper must consist in choosing such higher rules, norms, principles that guide both our individual actions and our social institutions.

Formal universality is not yet true universality, because it refers to contents (drives, needs) that are given rather than self-determined. The adequate determination of this still limited form of universality is "*self-determining universality, the will, or freedom*" (Hegel 1991: 52 [§21]). Freedom in the full sense must therefore be self-determination not only in form but also in content: "When the will has universality, or itself as infinite form, as its content, object [*Gegenstand*], and end, it is free not only *in itself* but also *for itself*—it is the Idea in its truth" (Hegel 1991: 52 [§21]). Truly free will has itself as its object. It is universality and indeed infinite

[18] Compare this to Raz's account of conflicting reasons (1975/1999: 25-26).

[19] This may be an interesting view for contemporary decision theory, which is based on the principle of utility maximization.

form. Therefore, the content itself must also be the infinite form. This remark explains free will in contrast to the first two forms.

The natural will is self-consciousness "external to itself" (Hegel 1991: 52 [§21R]). This exteriority of the natural will results from its sensibility. Since will implies self-consciousness, this sensibility is the exteriority of self-consciousness. Reflective will or volition is determined both by this sensibility and by the "thinking universality" (Hegel 1991: 52 [§21R]). The sensitivity is due to the fact that arbitrariness still has given contents as its object, the thinking universality consists in the fact that these are only possibilities and can be chosen or not chosen. The ability to choose implies the ability to abstract. Free will has itself as its object, and thus its own universality as its content, and not something merely naturally given. However, "*immediacy*" (Hegel 1991: 52 [§21R]) (natural will) and "*particularity*" (Hegel 1991: 52 [§21R]) (arbitrary will) are preserved in free will. Thus, free will, which has its own universality as its object, must not be misunderstood as consisting in the rejection and negation of sensually given contents. How these contents themselves are integrated, however, remains to be clarified. Nor is the reflection of arbitrariness simply negated. It, too, will continue to play a role in free will that remains to be determined.

Now the activity of thinking is precisely the process "by which the particular is superseded and raised to the universal" (Hegel 1991: 53 [§21R]). Thinking thus reasserts itself in the will, insofar as self-consciousness realizes itself through the will, leading to itself in the passage through various forms of will (natural, reflective): "Here is the *point at which it becomes clear* that it is only as *thinking* intelligence that the will is truly itself and free" (Hegel 1991: 53 [§21R]). The mind develops toward self-consciousness in the form of thinking and intelligence. In this respect, thinking is initially an epistemic attitude in the sense of a direction appropriate to the mind and the world. It is true that through this epistemic attitude the mind attains knowledge of itself. However, the mind does not fulfill its purpose of determining and realizing itself through mere thinking. Therefore, it must become active itself, which it can do only by transforming the mind-to-world direction of fit into a world-to-mind direction of fit, while retaining the first aspect. As Michael Quante observes:

> Self determination of the will has to be conceptualized as the logical interdependence of the first and second aspect which belong together in every person's free will. No person wills her "Ich" purely and without further content. And no person wants merely the content of her will. In wanting something a person always wants herself and the content as her content (Quante 1997: 65).

Hegel has three versions of a concept of will. Although the most minimal version of the will includes freedom, it lacks certain elements of the other forms. Central to Hegel's concept of free will in all its forms is a combination of determinism and indeterminism. Insofar as free will consists in self-determination, an undetermined will would not be free. On the other hand, there are forms of determinism that render a will unfree or not fully free. A free willing being is free to the extent that it determines itself. That it determines itself means that its determination is up to itself. That it is up to itself means that it is not determined by anything other than itself.

But that the determination of *x*'s will is up to *x*'s self is ambiguous: (1) it could mean that the *decision* for a determined content is up to *x*, or (2) it could mean that

the *content* for the decision is itself constituted by *x*. Even if arbitrary will is free in the first sense, it is not free in the second. Only if the determination of the content itself is up to a willing being can it be called free in the full sense. The only way that the content itself is subject to a willing being is that the content is nothing other than what the free willing being is—hence Hegel's expression "*the free will which wills the free will*" (Hegel 1991: 57 [§27]).

How can the content be understood if it is free will itself? As already indicated, my proposal is to say that the content of free will is rules.[20] Rules are free will itself, because they are action-guiding tools that are universal in the sense that they hold:

1. For *all* (or at least many different) *individuals*
2. For *all* (or at least many different) *times*
3. For *all* (or at least many different) *situations*

Of course, this proposal says nothing about which rules are legitimate or rational.[21] Understood in this way, free will has itself as its object insofar as it has rules as its objects, while these rules are themselves made by free will and thus by the content that is itself made by the will. To sum up Hegel's threefold analysis of the concept of free will:

Natural will	The most basic core ability to decide and begin to act (*acting freely*).
Arbitrary Will	The ability to think about given contents of will and to choose one over the other and to start acting on that choice (*particular autonomy*).
Free will	The ability to make up rules (second-order) of how to decide for contents (first-order), and to decide for such rules based on adequate self-understanding as a spiritual being (*universal autonomy*).

A free will skeptic might take all this for granted, but point out that it all depends on the claim that we have the ability to act freely in the first place. And this capacity itself may be incompatible with a deterministic world. So even Hegel's account stands or falls with his response to the traditional problem of free will and determinism. That we can, in principle, act freely is precisely what the *Science of Logic* is supposed to prove—among many other things, I am sure. The 'logical' underpinning of Hegel's understanding of free will that I have reconstructed so far is what I will call his metaphysics of free will.

4. Hegel's Metaphysics of Free Will

According to the proposed fourth criterion, a theory of free will must deal with the question of the compatibility of free will and causal determination. In order to see how Hegel deals with this, it seems quite natural to look at the very passage in Hegel's system that deals explicitly with causality. But this might seem unpromising, for, as Quante argues, in the introduction to the *Philosophy of Right*

[20] Note that this is not a second-order account of volition, for the truly free will has rules and institutions as its content, not first-order desires. See Quante 1997 for a very helpful and insightful discussion of Hegel within the second-order volition debate.
[21] Hegel answers this question in the three parts of the *Elements of the Philosophy of Right*.

Hegel assumes that causality is overcome in teleology (1997: 58).[22] If teleology, which Hegel discusses in the third book of the *Science of Logic*, overcomes causality and thus proves the inability of causality to explain certain phenomena, then why look at causality at all? (Quante 1997: 58). Although I concede Quante's and also Vieweg's claim, I read Hegel's chapter on causality (Hegel 2010b: 492-503) as relying on the concept of *causa sui* to prove how freedom is possible in a causally determined world.[23] In my reading, this depends on a much broader impetus of Hegel's *Science of Logic*: one problem it is supposed to solve is the possibility of freedom.[24] Instead of thinking about what free will should look like if it is to be compatible with the total determinism of the natural world, Hegel arranged the whole theory of determinism and the natural world in such a way as to ensure absolute free self-realization.[25] A guiding question of the *Science of Logic*, and perhaps of the whole system, would then be: How should the world be organized so that freedom can be realized in it?

The chapter on causality is part of the second book of the *Science of Logic*, the *Doctrine of Essence*. As Hegel makes very clear in the encyclopedic version, "the cause is thus in and for itself *causa sui*" (Hegel 2010a: 226 [§153R]). In the *Doctrine of Essence*, Hegel first analyzes the concept of causa sui, applies it to finite entities and shows that they realize the concept of causa sui only inadequately, and finally introduces a first stable phenomenon of causa sui within finite determinations, namely action and reaction, which leads to mutual causation or reciprocity of action (*Wechselwirkung*). Interestingly, mutual causation, as the very last category of the *Doctrine of Essence*, allows for the introduction of the term 'freedom'.[26] The chapter on causality shows not only that even for Hegel typical, finite causal processes (such as a rolling billiard ball hitting another and causing it to roll) are forms of self-causation. It also argues for the much more important phenomenon of substance causation, such that a substance exerts causal power out of itself.

Nevertheless, even for Hegel, a stone does not exercise substance causation, but only causally contributes to external causes acting upon it. This raises the question of which phenomena are real instantiations of the concept of substance

[22] Vieweg (2020: 61) also argues that the logic of the concept is decisive.

[23] For a thorough and detailed analysis of the chapter on causality, see my Meyer 2017. This talk of "causally determined" still leaves open the question of determinism as the claim that everything that happens is exclusively causally necessary.

[24] According to Pippin (1999: 195), the whole system can be understood as dealing with the question of the possibility and realization of freedom.

[25] I rely on the work of Christopher Yeomans (2012), who explicitly analyzes Hegel's formal, structural elements for freedom in the *Science of Logic*. I therefore side with Yeomans (2012: 4) here, albeit from a different angle: Hegel's theory of causality in the *Doctrine of Essence*. That Hegel arranged the whole theory of determinism and the natural world in such a way as to ensure absolute free self-realization can be explained by a core theme of German Idealism, which is to see subjectivity as the core phenomenon of all reality. If all reality can only be adequately understood as a form of realization of subjectivity, then the natural world must also be such that it allows for free self-realization, especially for those beings who are subjects par excellence.

[26] As Hegel writes: "Causality has thus returned to its absolute concept and has at the same time attained the concept itself. At first it is real necessity, absolute self-identity, in which the difference between it and the determinations which refer to each other in it are substances, free actualities, in relation to each other. Necessity is thus inner identity; causality is its manifestation, in which its reflexive splendour of substantial otherness has been sublated, and necessity is elevated to freedom" (Hegel 2010b: 504).

causation. A remark in the chapter on causality gives us a clue. Hegel warns against analyzing living beings (natural as well as spiritual) in terms of finite causality (that is, the deficient form of causality) (Hegel 2010b: 496). Regarding the causal influence on living beings, Hegel states that "anything that has an effect on a living thing is independently determined, altered, and transmuted by the latter, *for the living thing will not let the cause come to its effect*, that is, it sublates it as cause" (Hegel 2010b: 496).[27] Although this passage is not directly about substance causation, it is interesting because it contains an observation about a causal influence on living beings. The idea of determinism and its threat to the freedom of the will presupposes that willing beings are such that causal influences on them cause them to do things, in the same sense that a rolling billiard ball causes another to start rolling when the second is hit by the first. But living beings are not only capable of "sublating" external causal influences, they are also capable of being *causae sui*, of causing themselves to do certain things simply by beginning to do them.[28]

Hegel analyzes the categories that structure the world, such as causality, in such a way as to ensure the realization of freedom by arguing that they themselves are already a (albeit inadequate) form of the realization of freedom.[29] But does this mean that Hegel takes a compatibilist position and at the same time claims determinism? Robert Pippin ascribes to Hegel a form of compatibilism.

My proposal will be to say that for Hegel, freedom of the will is dependent on, and thus compatible with, a certain form of determinism, but incompatible with another. In the end, the position ends up as a form of libertarianism that allows for determination in the world even as it rejects the absolute determinism of everything that happens. In my reading, then, Hegel must be understood as an incompatibilist and indeterminist.[30]

5. Hegelian Libertarianism: A Systematic Proposal

How does it all fit together? In order to locate Hegel's understanding of free will within the traditional debate about free will and its compatibility with determinism, it is worth starting with his metaphysics of free will.

Free will is a form of self-determination. Causality as a category is also analyzed by Hegel as a form of self-determination. That is, his concept of finite causality, which consists in a regular process of quantity transfer between substances, is itself a (in principle) deficient form of self-determination. Free will is a form of self-realization that is not deficient in principle. Even if causal processes are determined in such a way that the instantiation of an event *B* is necessitated by an antecedent event *A*, this kind of determinism does not exclude the possibility of genuine self-determination, that is, a form of determination that is brought about by the very thing whose determination it is. This means that for Hegel not all

[27] For a thorough interpretation of this passage, see my Meyer 2023.
[28] Pippin, I think, therefore misinterprets the passage as evidence that Hegel "does not understand the possibility of freedom as depending on the possession by an individual of some kind of causal power, the power to initiate action by an act of will in some way independent of antecedent causal conditions" (Pippin 1999: 194).
[29] This follows from Hegel's claim that all forms that are philosophically analyzed must be understood as realizations of the Absolute, whereas the Absolute has the structure of subjectivity. Therefore, criterion C4 is fulfilled by the definition of Hegel's philosophy.
[30] I therefore disagree with both Pippin 1999 and Huesca Ramón 2015.

events are brought about by external necessity. What he does hold, however, is that many kinds of events must be such that they are regularly brought about by external causes. Let us call this $\text{Determinism}_{\text{weak}}$: Many events *B* in the world are such that their occurrence was caused by a prior external event *A*, and that every instantiation of *A* is followed, *ceteris paribus*, by an instantiation of *B*.

As an existential proposition, it is compatible with the claim that there are some events in the world that are not determined in this sense. In order for free will to be possible, $\text{Determinism}_{\text{weak}}$ must be true. For suppose it were false. Then there would be no event type, or at least there were not sufficiently many, in the world such that any instantiation of it would be regularly determined by the instantiation of some other earlier event type. Insofar as free will is possible only if there are some exercises of free will, and insofar as an exercise of free will consists in determining one's will at t_1 to make it the case that *p* at t_{1+n}, it must be possible to know that and how one could make it the case that *p* at t_{1+n}. It is only possible to know this if the world behaves in a regular way. If $\text{Determinism}_{\text{weak}}$ is false, the world is not regular. Therefore, $\text{Determinism}_{\text{weak}}$ must be true if free will is to be possible.[31] Thus, although $\text{Determinism}_{\text{weak}}$ must be true for the freedom of the will, a stronger version of determinism would make freedom impossible, which we can call $\text{Determinism}_{\text{strong}}$: All events *X* in the world are such that their occurrence was caused by some prior external event *Y*, and that every instantiation of *Y* is followed, *ceteris paribus*, by an instantiation of *X*.[32] There are real events in the world that are undetermined in this sense (human actions). But this does not preclude these events from being determined in any sense.[33] For the core concept of free will is that of self-determination, and self-determination is *a fortiori* determination. In order for self-determination to be possible, there must be real cases of substance or agent causation in the world—that is, the power to behave and thus to cause an event from itself.[34]

In the *Philosophy of Right*, Hegel distinguishes three different types of will: natural will (the ability to act freely), arbitrary will (the ability of particular autonomy), and free will (the ability of universal autonomy). All three presuppose a freedom of will that is compatible with and dependent on $\text{Determinism}_{\text{weak}}$. Only if the content of willing is free will itself can the will be called free in the full sense. For only then is the will capable of determining the content of willing itself, instead of merely choosing between externally given contents of willing. An act of free will in this sense consists in willingly following (knowing and valuing) social rules, which are themselves such that they are realizations of free will (see

[31] This could also be seen as an answer to the luck problem for indeterminists.

[32] I don't see Hegel taking a compatibilist position in the sense that free will is compatible with $\text{determinism}_{\text{strong}}$. Again, Pippin seems indecisive when he writes: "This means that the 'inappropriateness' of the principles of nature for explaining certain processes and alterations will have to be sought elsewhere in Hegel, not by appeal to and defense of this strong or voluntarist notion of spontaneity" (Pippin 1999: 195). Pippin seems to think that Hegel strongly rejected determinism somewhere in his system, which would make him an indeterminist and a libertarian given his position that free will is real.

[33] Of course, unless one accepts only events as (causally) determining factors for events.

[34] For a more recent version of agent causation, see Mayr 2011.

Neuhouser 2000: 82-113). Thus, free will and phenomena of free will take place on two different levels—subjective and objective freedom.[35]

The three spheres of freedom (abstract right, morality, and ethical life) are different institutional arrangements that realize objective freedom to different degrees. Individual freedom is realized by acting in accordance with these institutions. Here it becomes clear why, for Hegel, a mere individual will is not the highest and most important form of free will. Institutional arrangements remain robust no matter how one thinks about them and no matter how one values them. Although institutions depend on the basic form of individual free will for their emergence and reproduction, they themselves determine the wills and actions of the individuals who live within them. Only to a small extent are we as individuals ourselves agents of change of these institutions.

If one wanted to label Hegel's position, one could call him a libertarian (free will is incompatible with determinism$_{strong}$, but free will is real and therefore determinism$_{strong}$ is false, even though determinism$_{weak}$ must be true) who claims that our will is free in principle. He is an agent-causalist, whereas self-causation is to be understood as self-determination, and self-determination consists in deciding on rules and institutions.

I do not go into the question of which concrete rules and institutions are such that they actually realize freedom; that would be part of a reconstruction of Hegel's normative theory. What can be said, however, is that whatever rules are said to be rational, they must satisfy the criterion of constituting free will in the fullest sense.[36]

References

De Laurentiis, A. 2016, *Hegel and Metaphysics: On Logic and Ontology in the System*, Berlin: De Gruyter.

Dworkin, G. 1970, "Acting Freely", *Noûs*, 4, 367-83.

Gethmann, C.F. 1998, "Praktische Subjektivität und Spezies", in Hogrebe, W. (ed.), *Subjektivität*, Munich: Fink, 125-45.

Gethmann, C.F. 2016, "Was bleibt vom *fundamentum inconcussum* angesichts der modernen Naturwissenschaften vom Menschen?", in Quante, M. (ed.), *Geschichte—Gesellschaft—Geltung: XXIII. Deutscher Kongress für Philosophie, 28. September–2. Oktober 2014 an der Westfälischen Wilhelms-Universität Münster. Kolloquienbeiträge*, Hamburg: Meiner, 3-27.

Hegel, G.W.F. 1970/2004, *Hegel's Philosophy of Nature: Being Part Two of the "Encyclopaedia of the Philosophical Sciences" 1830) Translated [from the German] by Nicolin and*

[35] Neuhouser uses this difference to apply it to social freedom, which is the subject of "Ethical Life" (Hegel 1991: 189-380 [§§142-360]). I think that even "Abstract Right" (Hegel 1991: 67-132 [§§34-104]) and "Morality" (Hegel 1991: 135-86 [§§105-41]) are forms of social freedom, so that the distinction between subjective and objective freedom can also be applied to these spheres of freedom. Since I am not arguing for this view in this essay, I will omit the term 'social'.

[36] I would like to thank two anonymous reviewers for helpful comments and criticism. I also had the opportunity to present and discuss this article with several people. I would like to thank Roland Krause, Giovanna Luciano, Tobias Rosefeldt, and Myriam Stihl for helpful feedback and criticism.

Pöggeler's Edition 1959) and from the Zusätze in Michelet's Text 1847), Miller, A.V. (ed.), Reissue, Oxford: Clarendon Press.

Hegel, G. W.F. 1991, *Elements of the Philosophy of Right*, Wood, A.W. (ed.) and Nisbet, H.B. (trans.), Cambridge: Cambridge University Press.

Hegel, G.W.F. 2007, *Hegel's Philosophy of Mind*, Wallace, W. and Miller, A.V. (trans.), Oxford: Clarendon Press.

Hegel, G.W.F. 2010a, *Encyclopedia of the Philosophical Sciences in Basic Outline. Part I: Science of Logic*, Brinkmann, K. and Dahlstrom, D.O. (trans. and eds.), Cambridge: Cambridge University Press.

Hegel, G.W.F., 2010b, *The Science of Logic*, Di Giovanni, G. (trans. and ed.), Cambridge: Cambridge University Press.

Huesca Ramón, F. 2015, "Hegel's Concept of the Free Will: Towards a Redefinition of an Old Question", *Journal of Cognition and Neuroethics*, 3, 1, 309-25.

Illetterati, L. and Miolli, G. (eds.) 2022, *The Relevance of Hegel's Concept of Philosophy: From Classical German Philosophy to Contemporary Metaphilosophy*, London: Bloomsbury Academic.

Lewis, D. 1973, *Counterfactuals*, Oxford: Blackwell.

Mayr, E. 2011, *Understanding Human Agency*, Oxford: Oxford University Press.

Meyer, T. 2017, "Hegels wesenslogisches Kausalitätskapitel als Identitätstheorie der Kausalität", *Hegel-Studien*, 51, 91-119.

Meyer, T. 2020, *Verantwortung und Verursachung: Eine moral- und rechtsphilosophische Studie zu Hegel*, Hamburg: Meiner.

Meyer, T. 2023, "'Inadmissible Application': Some Notes on Causality and Life in Hegel", in Corti, L., Schülein, J.-G. (eds.), *Life, Organisms, and Human Nature: New Perspectives on Classical German Philosophy*, 59-77.

Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. 2013, "The Phenomenology of Free Will", in Russell, P. and Deery, O. (eds.), *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*, Oxford: Oxford University Press, 486-505.

Neuhouser, F. 2000, *Foundations of Hegel's Social Theory: Actualizing Freedom*, Cambridge (MA): Harvard University Press.

Pippin, R. 1999, "Naturalness and Mindedness: Hegel's Compatibilism", *European Journal of Philosophy*, 7, 2, 194-212.

Quante, M. 1997, "Personal Autonomy and the Structure of the Will", in Kotkavirta, J. (ed.), *Right, Morality, Ethical Life: Studies in G. W. F. Hegel's Philosophy of Right*, Jyväskylä: University of Jyväskyla, 45-74.

Raz, J. 1975/1999, *Practical Reason and Norms*, Oxford: Oxford University Press.

Rorty, R. 1999, "Comment on Robert Pippin's 'Naturalness and Mindedness: Hegel's Compatibilism'", *European Journal of Philosophy,* 7, 2, 213-16.

Ryle, G. 1949/2009, *The Concept of Mind*, London: Routledge.

Strawson, P.F. 1962/2008, "Freedom and Resentment", in Strawson, P.F. 2008, *Freedom and Resentment and Other Essays*, New York: Routledge, 1-28.

Vieweg, K. 2020, *The Idealism of Freedom: For a Hegelian Turn in Philosophy*, Leiden/Boston.

Williams, C. 1980, *Free Will and Determinism: A Dialogue*, Indianapolis: Hackett.

Wittgenstein, L. [1953] 2009, *Philosophische Untersuchungen / Philosophical Investigations*, G.E.M. Anscombe, P.M.S. Hacker, and J. Schulte (trans.), Revised fourth edition by P.M.S. Hacker and J. Schulte, Malden: Wiley-Blackwell.

Wolff, M. 1992, *Das Körper-Seele-Problem: Kommentar zu Hegel, Enzyklopädie (1830),* §389, Frankfurt: Klostermann.

Yeomans, C. 2012, *Freedom and Reflection: Hegel and the Logic of Agency*, Oxford: Oxford University Press.

Free Will: A Pseudo-Problem? Schlick on a Longstanding Metaphysical and Ethical Debate

Sofia Bonicalzi

Roma Tre University

Abstract

Free will, famously described by David Hume as "the most contentious question of metaphysics, the most contentious science", has long been a subject of intense debate, particularly regarding its compatibility with a deterministic universe and its implications for ethical questions, notably moral responsibility. Moritz Schlick, a leading figure in the Vienna Circle and the neopositivist movement, challenges the validity of this debate, asserting that it arises from linguistic and semantic confusions surrounding terms like 'freedom', 'determinism', and 'will'. Reflecting the neopositivist disdain for metaphysics and normative ethics, Schlick posits that once these concepts are properly defined, the problem of free will dissolves, though addressing it may be necessary when discussing moral and legal responsibility. This paper aims to elucidate Schlick's perspective on free will and responsibility, placing it within the early 20th-century debate, and highlighting his efforts to reconcile these concepts with the principles of physics while avoiding theoretical ambiguities.

Keywords: Free will, Laws of nature, Moral responsibility, Moritz Schlick, Role of ethics.

1. Introduction

Free will—described by David Hume (1975: 95) as "the most contentious question of metaphysics, the most contentious science"—has been the subject of a longstanding debate. As those familiar with metaphysical and ethical controversies know, the debate centers on whether free will is compatible with a universe governed by deterministic laws, thereby grounding moral responsibility for action. However, not everyone agrees that this is a genuine philosophical issue. According to Moritz Schlick, a key figure in the Vienna Circle and the neopositivist movement (Stadler 2001; Uebel and Limbeck-Lilienau 2022), the question of free will—despite its prominence in the Western philosophical tradition—is a false problem. It arises from linguistic and semantic misunderstandings, and overall vagueness surrounding concepts such as 'freedom', 'determinism', and 'will'.

Argumenta 11, 1 (2025): 129–145
ISSN 2465-2334

DOI 10.14275/2465-2334/20240.bon
First published: 9 July 2024

Once the correct meanings of these concepts are fully clarified, the problem would dissolve rather than being solved.

In fact, as indicated by the title of a section (*The Pseudo-Problem of Freedom of the Will*) in the seventh chapter (*When is a Man Responsible?*) of his *Problems of Ethics* (*Fragen der Ethik*) (1939),[1] Schlick shares the typical neopositivist aversion to traditional metaphysical problems, which he views as centuries-old sandcastles. In the same period, Carnap famously quips, after all, that metaphysicians are comparable to musicians without musical talent (1931), while Wittgenstein argues that problems related to the world as a totality are meaningless, as the world itself is not a fact but a collection of facts manifested in language (1922: 2.04, 5.6).[2]

As an heir and innovator of a tradition that includes Hume among its modern predecessors, Schlick thus expresses a natural skepticism towards a *vexata quaestio* that remains unresolved because it is fundamentally ill-posed: the sheer volume of discourse on this topic over the centuries is indeed, in Schlick's view, "one of the greatest scandals of philosophy" (1939: 143).

As the general title of the chapter suggests, the real target of Schlick's discussion of free will, which is described as a necessary but somewhat unwelcome part of the conversation, is the concept of responsibility—still intricately entangled with free will in contemporary debates (Bonicalzi 2019):

> The concept of responsibility constitutes our theme, and if in the process of its clarification I also must speak of the concept of freedom I shall, of course, say only what others have already said better; consoling myself with the thought that in this way alone can anything be done to put an end at last to that scandal (Schlick 1939: 144).

In this respect, Schlick's inquiry aims to elucidate the nexus between determinism and responsibility (both moral and legal) while crafting a theoretical framework that strives to attend to the subtleties of language, align with the principles of physics, and remain impervious to theoretical ambiguities. However, while briefly discussing early 20th-century physics advancements, the arguments presented by Schlick more significantly echo those employed by the British empiricist thinkers of the 17th and 18th centuries, with a particular emphasis on Hume, as one might expect.

Although many of the theoretical principles that inspired the philosophy of the Vienna Circle have been abandoned, Schlick's discussion of free will and responsibility remains a model of conceptual clarity. As we will see, the book also lays the groundwork for ongoing discussions on topics such as the role of ethics and the justification (consequentialist versus retributivist) of punishment. The aim of this paper is to elucidate Schlick's views on free will and responsibility, contextualizing them within the related debates of the early 20th century and the framework of his *Problems of Ethics*. Additionally, the present contribution will highlight some of the intersections of Schlick's ideas with contemporary debates on these topics.

[1] The quotes are from the American edition of the *Fragen der Ethik* (1930), published in New York in 1939, with the title *Problems of Ethics*.

[2] The quotes are from the British edition of the *Logisch-Philosophische Abhandlung* (1921), published in London in 1922, with the title *Tractatus Logico-Philosophicus*.

2. The Role of Ethics

From the outset of the *Problems of Ethics*, Schlick makes it clear that just as the general task of philosophy is to clarify the content of scientific propositions rather than to become a science itself, the primary aim of ethics should be to uncover and illustrate the meaning of moral judgments, thereby helping to explain human behavior from a psychological perspective—in a spirit that both methodologically and substantively aligns with current trends in philosophical and empirical moral psychology (Doris 2010).

However, while moral psychology is now typically regarded, and rightly so, as just one branch of ethics, Schlick adopts a more radical position, asserting that the *main* task of ethics as a discipline is to provide an answer to the question "'Why does man act morally?'" (1939: 30):[3]

> And here lies the proper task of ethics. Here are the remarkable facts which excite philosophic wonder, and whose explanation has always been the final goal of ethical inquiry. That man actually approves of certain actions, declares certain dispositions to be 'good', appears not at all self-explanatory to the philosopher, but often very astonishing, and he therefore asks his 'Why?' Now, in all of the natural sciences every explanation can be conceived as a causal explanation, a truth which we need not prove here; therefore the 'why' has the sense of a question concerning the cause of that psychical process in which man makes a valuation, establishes a moral claim (1939: 25).

In other words, ethics shall not be framed any further as producing a set of action-guiding normative principles (ethics, indeed, "creates neither the concept nor the objects which fall under the concept" (1939: 3) and the moral philosopher should not turn into a 'moralist' (1939: 2))—which would merely amount to unverifiable pseudo-propositions (1939: 2)—, but rather as a theoretical endeavor: "ethics is a system of knowledge and nothing else; its only goal is the truth" (1939: 1).

The main role of ethics is thus to elucidate and describe the necessary behavioral laws to which all entities are subject, assuming that they adhere to the principles of universal causality. It achieves this explanatory target by detailing the causal connections that govern human behavior, thereby aiming to function in a scientific manner. An example of such causal connections is provided by the so-called 'law of motivation', which states that when an individual is faced with different options, "the decision of the will proceeds in the direction of the most pleasant end-in-view" (1939: 38).[4]

This principle does not merely reflect a simplistic adherence to hedonistic values; rather, it seeks to mirror the fundamental mechanisms regulating behavior, including acts of self-sacrifice, where an individual might forgo their own immediate pleasure for the benefit of another, finding such a sacrifice more appealing. Schlick illustrates this with the case of a child who gives up a larger piece of

[3] While Schlick harbors doubts about normative ethics, his interests also encompass metaethical issues, including classic philosophical questions such as the meaning of 'moral' (1939: 79–99) or the existence of absolute values (1939: 100–119).

[4] Regarding the difference between laws and rules of behavior, see Schlick 1939: 41. To underscore Schlick's modernity, the notion that moral psychology serves to unveil the causal laws of motivation that underpin behavior remains pivotal in contemporary discourse. This is particularly evident in discussions exploring whether moral judgments and actions are driven by reasons or emotions (see e.g., Prinz 2008).

cake to offer it to another. In such cases, what superficially appears to be a less pleasant choice is linked to a range of emotionally charged mental states (such as thinking of a friend's disappointment if they were given the smaller piece). Therefore, opting for the seemingly less convenient choice may actually be—due to more nuanced factors beyond those captured by straightforward hedonistic considerations—the most pleasant choice, thus validating the law of motivation (1939: 42).

Challenging the Kantian division between the laws of nature and human freedom, Schlick thus establishes a more direct link between the realm of physics—the epitome of scientific rigor and empirical reliability—and the moral world as they are both governed by causal laws. Moral philosophers who attempt to isolate the moral from the natural realm, as if the two could be governed by distinct and irreconcilable systems of laws, are fundamentally mistaken, as this distinction lacks any valid foundation.

In this respect, it is noteworthy that while Schlick follows Wittgenstein's *Tractatus* in understanding philosophy as a clarificatory activity,[5] he diverges from him on the notion of 'laws of nature'. For Schlick, the laws of nature, as universal statements without precise references to space or time, cannot be directly empirically verified. However, their validity can be supported by the criterion of empirical verifiability. According to this criterion, a sentence expressing a law of nature must allow for the prediction of new data (Schlick 1948), including the future behavior of individuals in the case of human actions.

A notorious and significant limitation of Schlick's verificationism emerges here: a single empirical 'verification' or even multiple 'verifications' are insufficient to establish the universal validity of a law. Therefore, one should either refrain from considering the laws of nature as (universally valid) scientific statements or abandon the principle of verification altogether. Instead of being directly verifiable, the laws of nature may, however, serve as a model for individual statements (as rules through which such statements can be derived, primarily by scientists), which are more appropriately targeted by the verification process (Alai 1998: 120; Boniolo and Vidali 2003: 139–141).

For Wittgenstein, laws, regularities, and causal connections pertain instead to the domain of logic. Causality and the like—conceived of as the possibility of knowing future actions—do not pertain to the physical world, which is a totality of atomic facts: future facts cannot be inferred from present ones and the principle of causality has no specific role to play: "The events of the future cannot be inferred from those of the present. Superstition is the belief in the causal nexus" (Wittgenstein 1922: 5.1361) and "A necessity for one thing to happen because another has happened does not exist. There is only logical necessity" (Wittgenstein 1922: 6.37). In this respect, while the way we conceptualize the world is causally structured, this does not imply that the world is actually causally structured.[6]

[5] "The object of philosophy is the logical clarification of thoughts. Philosophy is not a theory but an activity. A philosophical work consists essentially of elucidations. The result of philosophy is not a number of philosophical propositions, but to make propositions clear. Philosophy should make clear and delimit sharply the thoughts which otherwise are, as it were, opaque and blurred" (Wittgenstein 1922: 4.112).

[6] This understanding of causality is reminiscent of Hume, but Wittgenstein might be closer to the Kantian idea of the necessity of a form that shapes and structures our knowledge (see Voltolini 2006).

In the second phase of his thought, Wittgenstein advances further in this direction, returning more poignantly to themes he had previously addressed only marginally. For Hume (1978), the experience of causal relations is needed to draw inferences about the nexus between causes and effects. For Wittgenstein, this experience is not even needed to the extent that the tendency to look for causal relationships is rooted in the spontaneous linguistic game of cause and effect. We instinctively react to a first event by considering it the cause of another event, without this tendency necessarily being supported by repeated experiences (1989; 2009; see Voltolini 2006).

Regarding the issue of free will, while Schlick maintains that the natural and human worlds must be homogeneous, Wittgenstein aligns more closely with Kant (compared to Hume) in supporting the distinction (*grammatical*, not real) between the two realms. In his *A Lecture on the Freedom of the Will* (1989), as in the *Tractatus*, Wittgenstein asserts that the causal model is just one among various types of explanations, while free will can be justified by referring to different systems. Both the determinist and the indeterminist models can be valid insofar as neither necessarily applies to the world, but both are foundational to its description: we can opt for the former when referring to the physical universe and for the latter when discussing the moral character of actions, which are subject to rewards and punishments. Depending on the grammar we choose, an action can be conceived of as free or unfree, as in the well-known example of the puppet, deceived into believing in its autonomy while being covertly directed:

> We talk of making decisions. Is there a case in which we would actually say that a man thought he decided, but actually didn't decide? In a prison you are normally locked in, said not to be free. I am in this room, free to go wherever I please. Suppose in the room below, there is a man, and he has certain people with him, and he says: 'Look, I can make W. go exactly where I want.' He has a mechanism, and he regulates it with a crank, and you see (with a mirror) that I walk exactly as the man wants me to. Then someone comes up to me and says 'Were you dragged about? Were you free?' I say: 'Of course. I was free'. Actually, there are cases which come pretty near to this. Man who could make someone choose the card he wanted him to choose. This is of course a primitive case. Everyone would say he chose freely, and everyone would say he made him choose what he wanted him to choose. People would say that the man in room above thought he was free, and actually did any damn thing people below wanted him to do (Wittgenstein 1989: 90).

The puppet can be described as free or unfree depending on the grammar we choose. This does not detract from the fact that, if we were to discover that all actions are predictable (as mentioned, in the *Tractatus*, freedom was defined as the impossibility of foreseeing future actions), we should abandon the grammar of freedom in favor of another one (Voltolini 2006).

3. Free Will as a Pseudo-Problem

The Schlickian chapter of the *Problems of Ethics* from which we began does not aim to prove the truth (or falsity) of determinism but rather to emphasize that the validity of any science must assume the principle of causation. This principle, equivalent to the existence of necessary laws, cannot be proven but must be assumed in the practices through which we organize the world. This is especially true if we consider the possibility of a science of human behavior in the form of

descriptive or explanatory ethics. Therefore, to explain moral behaviors, we must assume that psychological events are governed by the principle of causality, in the form of psychological laws.

In the early days of quantum mechanics—especially referring to the Copenhagen interpretation, which asserts its intrinsic indeterminacy—, it was specifically the principle of deterministic causation that was under scrutiny. In this context, Schlick argues that even indeterminacy at the quantum level does not preclude the acceptance of the causal principle but simply reduces the probability of making correct predictions: one must not fall into the fallacy of believing that freedom can be safeguarded through references to quantum indeterminacy, nor should one succumb to the erroneous dichotomy between freedom and deterministic causation. The misconception that determinism is incompatible with freedom primarily stems from a misunderstanding of terms such as 'freedom' and 'responsible'. This issue can be resolved by redefining these terms accurately. Hume (1975), indeed, already astutely observed that many metaphysical quandaries arise from the misuse of language, which has fueled centuries of philosophical debate.

If I may digress slightly, in more contemporary discourse, the issue of free will tends to be regarded as a profound philosophical problem rather than simply a matter of linguistic misunderstanding. Nevertheless, echoes of the Humean and Schlickian skepticism persist, cautioning us against falling into conceptual traps. For example, scholars like Chalmers (2011) argue that by leaving the definitions of free will and responsibility ambiguous, both compatibilism and incompatibilism risk being reduced to mere superficial labels. In fact, central to the debate between compatibilists and incompatibilists is their divergent interpretations of these concepts. This ambiguity suggests that discussions may continue without reaching substantive conclusions. For example, the question of whether free will and responsibility can coexist remains tangled in vague interpretations of these notions.

In this respect, following in the footsteps of the Scottish philosopher and influenced by the empirical spirit of the Vienna Circle, Schlick endeavors to achieve a clarity that eliminates obscure and misused metaphysical concepts from philosophical discourse, including notions like 'I' or 'consciousness'. Yet, it becomes evident to a contemporary reader that Schlick, while echoing early Wittgenstein in deeming the question of free will a pseudo-problem, does not entirely forsake metaphysics; instead, he proposes a solution that bears a striking resemblance to Hume's, so much so that he is often cited among those compatibilists enumerated by van Inwagen (1983: 13), and with good reason. Besides, Schlick acknowledges his debt to the Humean thought when he writes: "this pseudo-problem has long since been settled by the efforts of certain sensible persons; and, above all, the state of affairs just described has been often disclosed with exceptional clarity by Hume" (1939: 143).

Specifically, Schlick aims to dismantle the argument that if determinism holds, then all events, including human volition, are governed solely by the laws of nature, thus negating free will and responsibility. In essence, for those who could be termed, before the term was coined, 'incompatibilists', the concept of 'responsibility' presupposes the freedom of will: if individual choices were

inexorably determined by deterministic forces, notably the laws of nature, then our agency over our actions would be nullified.[7]

According to Schlick, the possible truth of determinism is no obstacle to the existence of free will. The incompatibilist stance harbors, in fact, a dual fallacy. Firstly, it misconstrues the essence of the concept of 'law of nature', erroneously conflating it with the notion of 'law' as it appears in legal statutes. Societies enforce compliance through sanctions, aiming to reconcile individual inclinations with legal statutes and suppress divergent behaviors. Conversely, within the scientific realm, the term 'law' denotes factual descriptions devoid of prescriptive mandates or coercive implications. In other words, it pertains to the realm of the 'is' rather than the 'ought'. Hume further underscores the perils of terminological ambiguity: the concept of 'necessity', particularly concerning the laws of nature, appears to insinuate the presence of force and compulsion. These elements, however, are imperceptible to us when we undertake actions that we consider free (Hume 1978).

If the conflation arises from the use of the same term, it is our usage of it in different linguistic contexts that makes it evident how we habitually, even if unconsciously, do not actually overlook this distinction. For instance, nobody would think that celestial mechanics dictate to planets the laws by which they must move; rather, they simply describe a certain state of affairs (or, we might say, construct a model of it).[8] Similarly, according to Schlick, the psychological laws that govern the will should be understood in the same vein, akin to the laws of nature rather than the civil laws. It should also be evident that individuals' natural desires are not the products of a coercive mechanism but function akin to the movement of planets—on a similar note, Dennett writes that "a jail without a jailer is not a jail" to illustrate that when individuals are determined by impersonal forces, true coercion does not exist (1984: 8).[9]

The tendency to anthropomorphize the natural laws fosters the second fallacy that Schlick seeks to rectify. This consists in the erroneous idea that universal validity implies irrefutable compulsion when, in truth, universality denotes the absence of exceptions rather than the presence of some form of compelling compulsion. Schlick's confidence in the power of the clarificatory process leads him to believe that if the distinction between these spheres were delineated, misunderstandings would simply evaporate. While this process is unlikely to suffice, as evidenced by the ongoing debate on free will and responsibility, matters are

[7] More precisely, Schlick's efficacious rendition of the standard incompatibilist argument can be paraphrased as follows: if determinism is true, meaning all events follow unchangeable laws, then my will is also determined by my innate character and motives. Consequently, my decisions are inevitable rather than free. If this is the case, I am not responsible for my actions because accountability would require the ability to influence my decisions, which I do not possess. My decisions necessarily arise from my character and motives, neither of which I control; my motives are external, and my character is shaped by inherent tendencies and external influences throughout my life. Therefore, determinism and moral responsibility are incompatible. Moral responsibility requires freedom, meaning exemption from causality (1939: 146).

[8] See also Wittgenstein 1989: 85: "To say that the natural law in some way compels the things to go as they do is in some way an absurdity".

[9] And it would be difficult not to mention here the famous Humean definition of freedom, which is "a power of acting or not acting, according to the determinations of the will; that is, if we choose to remain at rest, we may; if we choose to move, we also may". This kind of freedom is "universally allowed to belong to every one, who is not a prisoner and in chains" (1975: 8.23/95).

further complicated by the conflation of meanings extending to the antonyms—i.e., 'indeterminism' or 'acausality'—of such notions. True freedom, in its authentic sense, has nothing to do with the notion of 'exemption from the causal principle' or 'exemption from the laws of nature', which would entail liberation from causality and the natural laws (Schlick 1939: 149). Therefore, grounding freedom on immunity from such laws leads to the fallacious belief that proving indeterminism is necessary for preserving freedom. Conversely, the type of freedom essential for upholding human responsibility, which bridges theoretical deliberations on determinism or indeterminism with ethical considerations, operates within the realm of practical conduct, which is intricately entwined with moral issues.

While Hume made a pertinent distinction between 'indeterminism' and 'freedom' (1978), his usage of the term 'freedom' interchangeably for both the will and conduct (or action) inadvertently muddied the waters of the debate. According to the refined Humean perspective, which Schlick revisits, only freedom of conduct is unequivocally attributable to the agent, and this alone forms the basis for moral and legal responsibility. This freedom of conduct finds its logical antithesis in 'compulsion', which, if present, would nullify the agent's accountability.

However, the configuration of this freedom of conduct raises pertinent questions. One might wonder whether this notion adequately captures our conventional understanding of freedom, or if it appears so counterintuitive compared to common sense that it becomes uninteresting. Additionally, can freedom, conceived of as the mere absence of compulsion in the tradition of Hobbes and Hume, align with our intuitive moral and legal sensibilities regarding free will and responsibility, or must these intuitions be discarded as illusory?

From both moral and legal perspectives, as Schlick suggests, a free individual is one who encounters no impediments in realizing their desires. Besides, the notion that free will and accountability originate from an individual's capacity to act by their own volition is a recurring theme found in the works of numerous authors, ranging—just to mention a few well-known names—from Hobbes (Hobbes and Bramhall 1999) to Frankfurt (1971), representing a significant strand within contemporary compatibilism (Bonicalzi 2019). Schlick further highlights how the existence of intermediate scenarios between freedom and compulsion lends additional credence to this thesis: society typically deems someone responsible for a crime committed under drug influence if the drug was consumed voluntarily, whereas actions of individuals with mental deficiencies (internal compulsion) tend to be viewed as less culpable.

This leads, however, to a further inquiry, one which Schlick does not shy away from: what precisely is meant by 'responsibility'? In what circumstances, beyond those already outlined, can we legitimately assign responsibility to an agent? From a legal standpoint, the issue of responsibility is intricately linked with that of punishment. Indeed, the notion of current or historical responsibility, akin to 'accountability' as discussed by Gary Watson and "in which one individual or group is held by another to certain expectations or demands" (Watson 1996: 235), must be carefully determined in anticipation of future repercussions.

Schlick argues that a truly civilized society cannot find the justification for punishment in the primal desire for revenge against alleged wrongdoing. The notion that increasing pain can somehow result in something good through further infliction of pain is surrounded by an almost barbaric aura. Conversely, Schlick understands and justifies punishment as an educational tool that serves to guide

and shape the motivations underlying behavior, to prevent the repetition of an act (by the same agent) and encourage different behavioral determinations (in other agents). In advocating this perspective—consequentialist and forward-looking—, Schlick outlines an approach to punishment opposing standard retributive conceptions. This viewpoint is still upheld by those, who, on various grounds, defend analogously forward-looking justifications of punishment, valuing aspects such as reconciliation, societal security, and the offender's reintegration into society (Caruso and Pereboom 2022).[10]

When the focus shifts to the individual perpetrating the act, the pertinent question becomes: who is the correct target of punishment (or reward)? The challenge lies not in uncovering the initial causal factors of the act—whether it be the family or social environment—but in identifying the individual in whom the decisive conjunction of causes culminates. Schlick maintains that this problem is solved once an agent is identified as the appropriate locus for the motivations behind the action ("the one *upon whom the motive must have acted*" (1939: 152)), being an apt target of either punishment or reward. To strengthen the intuitiveness of his forward-looking take on responsibility and punishment, Schlick further contends that the reluctance to punish those afflicted by mental deficiencies stems from the futility of attempting to modify their conduct through promises or threats. Thus, blame is not directed at those under threat but at those who impose the threat, as it is their behavior that society seeks to influence and alter for future societal harmony.

A fascinating aspect of Schlick's perspective on responsibility revolves around the connection between one's subjective sense of responsibility and the concept of punishment. How does one consider the autonomous judgment that individuals pass upon themselves, feeling somehow deserving of punishment and censure? If punishment primarily serves an educational purpose, what significance does the sense of guilt hold for those who perceive themselves as having committed wrongful deeds? For proponents of the nexus between freedom and indeterminism, the individual's assessment of their actions stems from a subjective awareness of having acted freely, devoid of external compulsion, while spontaneously pursuing their innate desires.

The absence of external imposition and the capacity to choose otherwise in similar circumstances represent key aspects of free will and responsibility within the contemporary debate. Apart from the agent's *subjective* sense of acting freely, Schlick's stance on the agent possessing an *objective* 'capacity to do otherwise' closely mirrors that of George Edward Moore (1912). Moore notoriously writes that this capacity must be understood as meaning that the agent could have acted differently, but only if they had chosen to do so. Embracing Moore's interpretation, one views the requirement for the possibility of alternative actions as compatible with a deterministic universe, as the hypothetical conditionals would maintain their truth even if alternative choices were not actualized. Schlick, echoing Moore, argues that under identical conditions, the agent could have acted differently if alternative motives were present—namely if they desired something different.

[10] Indeed, the contemporary debate on punishment continues to be dominated by the opposing viewpoints of retributivism and consequentialism. Retributivists advocate for justice restoration through punishment of those deemed deserving, while consequentialists prioritize maximizing general utility (Altman 2021). However, proponents argue that a *mixed* theory of punishment, which integrates constraints from negative retribution with principles of act utilitarianism, provides a more nuanced and justified approach (Hart 1968).

The subjective feeling that one could have made a different decision under the same circumstances provides scant insight into the theoretical inquiry regarding the validity of the principle of causality, as individual consciousness is not germane to the validation of principles:

> How this indubitable experience ever came to be an argument in favor of indeterminism is incomprehensible to me. It is of course obvious that I should have acted differently had I willed something else; but the feeling never says that I could also have willed something else, even though this is true, if, that is, other motives had been present (Schlick 1939: 155).

Moreover, such a feeling or experience does not necessarily imply the absence of action-determining causes but rather encapsulates a subjective sense of freedom derived from the perception of being able to act by one's desires—desires that, as Hobbes previously noted (Hobbes and Bramhall 1999), are not inherently voluntary. In other words, the assertion 'I could have acted otherwise' conveys nothing beyond the notion that, given different motivations and in alignment with the laws of volition, alternative behavior could have ensued. In this respect, I assume responsibility for an action to the extent that my desires align with the motives that instigated it, i.e., I am the locus where the causes converge.

Schlick's aims to reframe the debate on free will, circumventing his adversaries by illustrating not only the compatibility of determinism with responsibility but also the impossibility of grounding responsibility otherwise; indeed, it could not be sustained in a universe deemed indeterministic. This argument, previously articulated by Hume and echoed in various forms throughout the 20th-century discourse (notoriously, by Hobart (1934)), posits that an indeterministic cosmos yields pure randomness, thus obliterating any semblance of responsibility.[11]

Moreover, from a pragmatic perspective, in an indeterministic universe, the presumed connection between responsibility and punishment would disintegrate. Since regulating an individual's conduct requires influencing their motivations through incentives and penalties, such regulation would become illogical if the agent's choices lack a determinable cause. Thus far, we have construed an agent as responsible insofar as their conduct aligns with discernible motivations. Indeed, if an individual were to possess freedom—defined as conduct devoid of underlying reasons—they might find themselves in a position similar to someone subjected to compulsion, as they would lack mastery over their actions, which would not arise from internal motivations.

Though the agent's self-perception in this regard offers scant theoretical insight, they nonetheless feel varying degrees of responsibility contingent upon the motives animating their behavior—a sentiment echoed by societal assessments,

[11] Regarding Schlick's view that there are no alternatives to the dichotomy between randomness and determinism, Popper, a longstanding critic of the Vienna Circle, states "Hume's and Schlick's ontological thesis that there cannot exist anything intermediate between chance and determinism seems to me not only highly dogmatic (not to say doctrinaire) but clearly absurd; and it is understandable only on the assumption that they believed in a complete determinism in which chance has no status except as a symptom of our ignorance. (But even then it seems to me absurd, for there is, clearly, something like partial knowledge, or partial ignorance.)" (1972: 227). Similarly, in rejecting the dichotomy between randomness and determinism, contemporary debates on free will often acknowledge the possibility of probabilistic forms of causation (Clarke 2003).

evident in the allocation of rewards and punishments based on motives and reasons. Given that no one in daily life challenges the validity of the principle of cause, there exists no excuse for conflating freedom with a sort of willful impulse devoid of causality. However, while no alternative to the dichotomy between causality and randomness exists, this does not imply that we can furnish evidence for determinism beyond the arguments presented. Rather, it underscores the impossibility of applying the concept of responsibility to conduct without presupposing the practical validity of the principle of cause.

4. An Early Libertarian Rebuttal of Schick's View on Free Will

Far from settling the dispute regarding causality and free will, Schlick's intervention sparked a robust rebuttal from philosophers who contended that the enduring quandaries of metaphysics could not be neatly resolved through mere analysis of semantic nuances but warranted, if anything, a reexamination in light of newly acquired theoretical tools. The crisis of logical neo-positivism, particularly the erosion of verificationism, served as a catalyst for reigniting discussions—both within the Anglo-Saxon sphere and beyond—around the conventional themes of metaphysics, ethics, and a plethora of questions transcending linguistic realms.

This trend is reflected in the contemporary state of the debate on free will and moral responsibility, which is further enriched by insights from the sciences of the mind (Bonicalzi and De Caro 2022). However, its roots extend back to the 1930s.

In this section, we will briefly examine an early rebuttal of Schlick's thesis. In 1938, C.A. Campbell, a scholar in Glasgow, published a volume, titled *In Defence of Free Will*, advocating not only the continuation of the debate on free will and responsibility but also a response aligned with traditional arguments favoring the incompatibility between determinism and freedom. In 1951, Campbell revisited the book's arguments with his article *Is Free Will a Pseudo-Problem?*, positioning himself—as indicated by the title—as a libertarian counterbalance to Schlick's theses.

According to Campbell, though the logical underpinnings of Viennese-style verificationism have long been scrutinized and largely discarded, the philosophical scandal lies in the persistence of the anti-metaphysical biases that once bolstered the neo-positivist tenets. In this respect, Campbell's critique goes beyond a mere repudiation of Schlick's logical premises, as he contends that an analysis of terms alone cannot suffice to resolve theoretical conundrums that, for thinkers like himself, transcend linguistic misunderstandings. While Campbell articulates a traditional defense of free will, his aim extends to a broader objective: refuting the reduction of metaphysical inquiries to pseudo-problems—a stance reminiscent also of Carnap's renowned *Scheinprobleme in der Philosophie* (1928).

Given this chosen battleground, the exploration of free will becomes an inevitable endpoint, particularly as Campbell regards it as the quintessential problem of philosophy—a sentiment shared by many who opposed the neo-positivists.[12] Campbell's critique of Schlick's text operates on two fronts: challenging its portrayal of free will as a pseudo-problem and rejecting its compatibilist stance, which is untenable to a libertarian thinker like himself. Campbell's method of refutation, however,

[12] "We do know of one traditional problem that is definitely on the black list of the *avant garde*—the problem of 'Free Will' […]. A plain obligation now lies upon philosophers who still believe that 'Free Will' is a genuine problem to explain just where, in their opinion, the case for the prosecution breaks down. To discharge this obligation is the main purpose of the present paper" (Campbell 2004: 18).

does not proceed from a denial of the logical assumptions of Schlick's chapter on free will and moral responsibility. It becomes clear that following its steps naturally leads, as we have experienced, to the same conclusions.

The Scottish philosopher's suggestion is rather to begin with a different question that identifies the underlying premise implicit from the outset in Schlick's theses (and explicitly stated already by Hobbes (1983)): does freedom consist solely in the absence of compulsion? In Campbell's interpretation, the most glaring fallacy of Schlick's argument lies in the fact that such a definition of freedom—as the absence of compulsion—fails to adequately account for the notion of responsibility, which implies something way beyond that.

According to Schlick, the question of attributing responsibility corresponds to that of assigning punishments, which serve a dual function of rehabilitation (towards the actor) and deterrence (towards others). Therefore, identifying the perpetrator of an act would simply consist in identifying whom to target with educational influence. As mentioned, what we seek is not the remote cause, which is deemed irrelevant in terms of punishment, since its contribution cannot be quantified and is beyond our reach anyway.

In Campbell's view, the implicit consequence of Schlick's position might be that an agent could be considered morally responsible only if we believe that the motives guiding their conduct are such that rewards and punishments could potentially influence them and serve an educational function. For instance, if a person acts under the control of an external agent, we cannot consider them responsible, nor can we expect our attempts to influence their behavior to succeed. From this perspective, the form of freedom that a libertarian would describe as 'contra-causal'—i.e., identifying the philosophical concept that an agent's actions are not wholly determined by preceding causes or factors external to the agent so that they have the capacity for genuine, undetermined choice or volition, independent of deterministic causality—appears to play no significant role.

As mentioned, Campbell finds untenable the reductionism that would diminish the issue of free will solely to a problem of terminological imprecision. Acknowledging that freedom entails a break in the causal chain (as in contra-causal freedom) has little to do with confusion between prescriptive and descriptive laws, but it is rather connected with the idea that an unbroken causal chain leads to further conclusions incompatible with moral responsibility.

The crux lies instead in what Moore (1912) had already highlighted and which Schlick himself acknowledges: the admission of freedom implies the possibility of doing otherwise. How should this latter concept be understood? Campbell believes that if we adopt the concept of responsibility in its pre-philosophical sense,[13] which also concerns us from the moral point of view, we arrive at assumptions radically opposed to those underlying Schlick's reasoning: in a libertarian perspective, common sense dictates that an agent is not considered morally responsible unless granted the possibility of doing otherwise (in the actual world), a condition that cannot be fulfilled without breaks in the causal chain. Schlick may argue that determinism is compatible with freedom only insofar as he accepts a pale image of freedom that does not align with the concept of responsibility that we ordinarily value and that plays a role from the moral point of view.

[13] Contrary to Campbell's suggestion, numerous scholars have recently highlighted that the pre-philosophical concept of free will is far less evidently aligned with libertarian principles (Nahmias et al. 2005).

Moreover, if punishment were solely presumed to have an educational and preventive function, as Schlick suggests, one might question why it would exclusively be imposed on agents deemed morally responsible. The same effect could also occur in those who are not responsible but could still benefit from the deterrent effect of punishment, as in the case of animals (whose behavior could be influenced by our actions). The real reason we do not consider an animal morally responsible lies instead in our understanding that it could not have acted otherwise, that its behavior is directed by a series of causes beyond its control, whereas responsibility implies a form of contra-causal freedom which Schlick seeks to deny.

While it is true, as Campbell notes, that for Schlick the meaning of punishment lies in its potential effect on the agent, this critique of Schlick's thesis is not particularly persuasive. What matters to the Austrian philosopher is not the actual ability to locate the culprit but rather the theoretical possibility of identifying the responsible agent as the endpoint where the determining causes of the action converge. This is a significant distinction, comparable to the difference between verificationism (according to which the truth value of a statement depends on the theoretical possibility of its empirical verification) and actual verification (which may be impossible due to our current cognitive-technological limitations) (Schlick 1938). After initial hesitation, Campbell himself appears more willing to credit Schlick, interpreting his opponent's statements in a less literal sense.

However, even assuming that Schlick only considers the theoretical applicability of punishment, a libertarian would find it difficult to accept his view: if we were to consider punishment solely in relation (even theoretically) to its effect, we should not accept the relevance, in the case of responsibility attributions, of (potentially excusing or exempting, in a Strawsonian fashion (Strawson 1962)) factors that the law typically considers when determining the severity of punishment, such as the role of the social environment or the type of education the culprit received. According to Campbell, if we are willing to follow Schlick in equating punishment with responsibility, it is solely because our concept of 'punishment' inherently includes a primal feeling of revenge, which however, according to Schlick, cannot legitimately be the foundation of our penal system.

Moreover, if one were to view punishment purely as an educational tool, as Schlick suggests, it ought to be disconnected from any association with the notion of 'moral blameworthiness' (a similar but converse argument would apply to moral desert and to associated feelings). Common sense indeed distinguishes between an agent considered morally blameworthy (because they voluntarily committed an action with full awareness of its consequences) and one who commits an action that should be avoided but is not inherently worthy of moral blame. For instance, imagine a man who, due to a misunderstanding and perhaps at great personal sacrifice, unwittingly performs an action deemed dangerous by the community. In this case, we might feel justified in punishing him, but we would not morally blame him for the mistake.

However, Campbell's argument lacks again complete persuasiveness as it assumes that Schlick's views preclude differentiation between (a) those consciously pursuing their desires and (b) those unwittingly engaging in perilous conduct due to misunderstanding. Nonetheless, in a classic compatibilist perspective, which rejects the notion of desert in favor of a consequentialist and forward-looking approach to responsibility and punishment, the distinction between (a) and (b) hinges on the specific causal history of reasons and motivations leading the agent to perform a particular act, rather than on the illusory autonomy of the subjects regarding the set

of determining causes. In this respect, Hobbes already observes that when we perceive that a force is driving us, we acknowledge it as a necessity; however, when we do not notice this force, we assume it does not exist and attribute our actions to freedom rather than to any causes (Hobbes and Bramhall 1999). Moreover, regarding the subjective sense of responsibility and guilt, Schlick clarifies that these can be elucidated through the agent's acquired awareness of acting on their motives, as opposed to being compelled by external forces (1939: 154).

Campbell further contends that the crux of Schlick's fallacy lies in exclusively tethering punishment and responsibility to the potential for influence on the agent, neglecting contra-causal freedom—a concept Schlick outright denies. Contrary to the claims of determinists, Campbell assumes that this kind of contra-causal freedom underpins the pre-philosophical concept of responsibility. Moral considerations, particularly the attribution of moral responsibility, necessitate holding an agent accountable for actions they could have avoided. As Schlick is aware of, the question of what it means to say that 'the agent could have done otherwise' is therefore crucial. Depending on the interpretation one gives to this notion, the reasoning diverges significantly. According to Campbell, when we hold an agent accountable because we believe they could have acted differently, the moral weight of our judgment rests on our implicit belief in the real (rather than the conditional) possibility of alternative actions—specifically, that the agent could transcend the causal chain. However, the critical issue remains: where does this causal chain break? This question serves as a crucial juncture in the debate between compatibilists and incompatibilists, highlighting significant variations among incompatibilist perspectives that advocate for such a break (Bonicalzi 2019).

To underscore the importance of considerations based on the availability of alternative possibilities, it is evident that we frequently excuse individuals whose actions appear to be significantly influenced by contextual or social pressures, acknowledging the impact these factors have on their behavior. In Campbell's libertarian view, when we speak of agent A and claim that in a given circumstance they could have acted differently, we simply mean that individual A, under the same circumstances and with the same natural desires, could have made a different decision and acted accordingly.[14] Hence the rejection of the solution proposed by Moore and accepted by Schlick, as well as of a similar one developed by Patrick Howard Nowell-Smith (A could have acted differently, but not under the same circumstances and with the same natural desires).[15]

For Campbell, the ability to translate desire into action—which, according to Schlick and the classic compatibilist tradition, constitutes the entirety of

[14] "What then does one mean in this class of cases by 'A could have acted otherwise'? I submit that the expression is taken in its simple, categorical meaning, without any suppressed 'if' clause to qualify it" (Campbell 2004: 32).

[15] "To say that Nero might have acted otherwise is to say that he could have decided to act otherwise and that he would have so decided if he had been of a different character. If Nero had been Seneca, for example, he would have preferred suicide to matricide. But what could 'If Nero had been Seneca . . .' possibly mean? Unfulfilled conditionals in which both terms are names of individuals constitute, admittedly, a thorny philosophical problem; but it is clear, I think, that if 'If Nero had been Seneca' means anything at all, it is a quasi-general proposition which can be analysed either as 'If Nero had had the character of Seneca' or 'If Seneca had been emperor' or in some similar fashion. None of these analyses are incompatible with the Determinist's contention that, as things stood, Nero could not have abstained" (Nowell-Smith 1948: 50).

freedom—represents only one aspect of true freedom. Schlick and Campbell would both agree that A is not morally responsible for missing an appointment if they break their leg an hour beforehand. However, as Campbell suggests, there are situations where the agent's ability to act according to their desires is not in question (consider, for instance, a case in which A voluntarily tells a lie). In such cases, if we believe the agent could have chosen otherwise, we automatically consider them morally blameworthy. In these contexts, it is insufficient to interpret 'A could have done otherwise' as 'A could have done otherwise, had they chosen to do otherwise,' as in the Moorean interpretation. To preserve the concept of responsibility, it is essential to acknowledge that if an agent cannot break the causal chain, then true freedom—and thus responsibility—cannot exist. The agent must have the genuine ability to choose differently in the given circumstances.

Many other philosophers will similarly highlight the inadequacies of the conditional analysis of the possibility to do otherwise, leading to its increasing abandonment (though see Clarke 2009). In addition to this specific critique, Campbell deserves recognition for effectively rehabilitating the concept of free will, particularly in light of its connections with moral issues, as a legitimate philosophical topic rather than a pseudo-problem. Nevertheless, compatibilism will remain one of the most popular positions on free will, albeit with contemporary views that present refinements reflecting the more advanced state of the debate (Bonicalzi 2019).

5. Final Notes

This article aims to discuss Schlick's stance on free will and responsibility within the early 20th-century landscape, particularly in the context of his long-distance debate with Campbell, while also referencing contemporary discussions on the topic. Several key aspects of Schlick's position have emerged, including his advocacy for descriptive over prescriptive ethics, the neopositivist assertion that metaphysical issues can be resolved by clarifying the meaning of contentious concepts, the acceptance of the Moorean analysis of the possibility to do otherwise, the noncoercive nature of natural laws, and the idea that responsibility and punishment serve a forward-looking purpose. In the subsequent decades, some of these ideas, such as the skepticism about metaphysics and normative ethics and the acceptance of the conditional analysis of the possibility to do otherwise, will be abandoned or profoundly revised, while others, like the forward-looking role of responsibility and punishment and the importance of a causal understanding of moral psychology, will continue to develop. In any case, Schlick's clear and direct approach to intricate moral questions solidifies the *Problems of Ethics* as a philosophical classic of enduring value.

References

Alai, M., 1998. *Filosofia della scienza del '900*. Roma: Armando Editore.

Altman, M.C., 2021. *A theory of legal punishment: deterrence, retribution, and the aims of the state*. Abingdon: Routledge.

Bonicalzi, S., 2019. *Rethinking moral responsibility*. Milan-London: Mimesis International.

Bonicalzi, S. and De Caro, M., 2022, How the Libet tradition can contribute to understanding human action rather than free will. *In:* C.J. Austin, A. Marmodoro, and A. Roselli, eds. *Powers, time and free will*. Cham: Springer, 199–225.

Boniolo, G. and Vidali, P. 2003. *Introduzione alla filosofia della scienza*. Milan: Mondadori.

Campbell, C.A., 1938. *In defence of free will*. London: Allen & Unwin. Reprinted in C.A. Campbell, 2004. *In defence of free will: with other philosophical essays*. London: Routledge.

Campbell, C.A., 1951. Is free will a pseudo-problem? *Mind*, 60 (240), 245–261. Reprinted in C.A. Campbell, 2004. *In defence of free will: with other philosophical essays*. London: Routledge, 17–34.

Carnap, R., 1928. *Scheinprobleme in der Philosophie*. Frankfurt a. M.: Suhrkamp.

Carnap, R., 1931. Überwindung der Metaphysik durch logische Analyse der Sprache. *Erkenntnis*, 2, 219–241.

Caruso, G.D., and Pereboom, D., 2022. *Moral responsibility reconsidered*. Cambridge, MA: Cambridge University Press.

Chalmers, D., 2011. Verbal disputes. *Philosophical review*, 120 (4), 515–566.

Clarke, R., 2009. Dispositions, abilities to act, and free will: the new dispositionalism. *Mind*, 118 (470): 323–351.

Clarke, R., 2003. *Libertarian accounts of free will*. New York: Oxford University Press.

Dennett, D.C., 1984. *Elbow room: the varieties of free will worth wanting*. Cambridge, MA: MIT Press.

Doris, J.M., ed., 2010. *The moral psychology handbook*. Oxford: Oxford University Press.

Frankfurt, H.G., 1971. Freedom of the will and the concept of a person. *The journal of philosophy*, 68 (1), 5–20.

Hart, H.L.A., 1968. *Punishment and responsibility*. Oxford: Oxford University Press.

Hobart, R.E., 1934. Free will as involving determination and inconceivable without it. *Mind, New Series*, 43 (169), 1–27.

Hobbes, T., 1983. *De cive. The English version* (entitled, in the first edition, *Philosophical Rudiments concerning government and society*). H. Warrender, ed. Oxford: Clarendon Press.

Hobbes, T. and Bramhall, J., 1999. *Hobbes and Bramhall on liberty and necessity*. V. Chappell, ed. Cambridge, MA: Cambridge University Press.

Hume, D., 1975. *Enquiries concerning human understanding and concerning the principles of morals*. L.A. Selby-Bigge and P.H. Nidditch, eds. Oxford: Clarendon Press.

Hume, D., 1978. *A treatise of human nature*. L.A. Selby-Bigge and P.H. Nidditch, eds. Oxford: Clarendon Press.

Moore, G.E., 1912. *Ethics*. London: Williams and Norgate.

Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J., 2005. Surveying freedom: folk intuitions about free will and moral responsibility. *Philosophical psychology*, 18 (5), 561–584.

Nowell-Smith, P.H., 1948. Freewill and moral responsibility. *Mind*, 57 (225), 45–61.

Popper, K.R., 1972. *Objective knowledge: an evolutionary approach*. Oxford: Clarendon Press.

Prinz, J., 2008. *The emotional construction of morals*. Oxford: Oxford University Press.

Schlick, M., 1938. *Form and content: an introduction to philosophical thinking*. *In*: M. Schlick, *Gesammelte Aufsätze 1926-1936*. F. Waismann, ed. Vienna: Gerold.

Schlick, M., 1939. *Problems of ethics*. D. Rynin, transl. New York: Dover Publications.

Schlick, M., 1948. Positivism and realism. *Synthese*, 7 (1), 478–505.

Stadler, F., 2001, *The Vienna circle*. Wien: Springer.

Strawson, P.F., 1962. Freedom and resentment. *Proceedings of the British Academy*, 48: 187–211.

Uebel, T. and Limbeck-Lilienau, C., eds., 2022. *The Routledge handbook of logical empiricism*. London: Routledge.

Van Inwagen, P., 1983. *An essay on free will*. New York: Oxford University Press.

Voltolini, A., 2006. Introduzione: Wittgenstein tra causalità e libertà. *In*: L. Wittgenstein, *Causa ed effetto seguito da Lezioni sulla libertà del volere*. Torino: Einaudi, VII–XXVI.

Watson, G., 1996. Two faces of responsibility. *Philosophical Topics*, 24 (2), 227–248.

Wittgenstein, L., 1922. *Tractatus logico-philosophicus*. C.K. Ogden, transl. London: Routledge & Kegan Paul.

Wittgenstein, L., 1989. A lecture on freedom of the will (Notes by Yorick Smythies at the Lecture). *Philosophical investigations*, 12 (2), 85-100.

Wittgenstein, L., 2009. *Philosophical investigations*. P.M.S. Hacker, ed. Oxford: Blackwell.

Advisory Board

Intra-Personal Compromises

Juha Räikkä

University of Turku

Abstract

The most usual philosophical questions about compromises have been those related to *inter-personal* compromises, in which parties are compromising with each other, rather than *intra-personal* compromises, which are often psychologically demanding. This paper aims to fill the gap in the discussion and briefly analyze the nature of intra-personal compromises. The starting point here is the assumption that inter-personal compromises cannot be made without intra-personal compromises, although intra-personal compromises are common even when they are not linked to inter-personal compromises. The main question addressed in the paper is whether the intra-personal compromises that we accept in all kinds of contexts are similar to those intra-personal compromises that we make when we compromise with others. I argue that they are more or less similar, although there are also some distinctive features in intra-personal compromises that are involved in inter-personal compromises. When a person makes an intra-personal compromise in the context of an inter-personal compromise she is forced to act under uncertainty, as she cannot know beforehand what options are really available. The price of the compromise is known only after the negotiation process. This is a special feature, or so I will claim.

Keywords: Inter-personal compromise, Intra-personal compromise, Bargaining, Uncertainty.

1. Introduction

Philosophical debate about compromises has concerned a variety of topics. Among other things, it has been asked whether compromises should be fair, what kinds of compromises are morally unacceptable and demonstrate lack of integrity, are compromises suitable for both conflicts of interests and conflicts of principles, what does it require to reach a compromise, and should we emphasize the processes rather than the results of compromises when evaluating them (Braybrooke 1982; Ceva 2016; Hall 2022; Huxtable 2014; Luban 1985). What is common to these and similar questions is that they all concern *inter-personal* compromises, that is, compromises in which parties compromise with each other. The most common philosophical questions about compromises have not been about *intra-personal*

Argumenta 11, 1 (2025): 149–162
ISSN 2465-2334

DOI 10.14275/465-2334/20240.rai
First published: 10 May 2024

compromises. A person may have to choose between her principles or values, and she may feel compromised by something after the choice. Intra-personal compromises are often unpleasant and difficult, but issues of a person's inner life have not been the focus of the discussion. Philosophers have rather been interested in the logic and features of inter-personal compromise.

There are some exceptions. Some authors have written about the regret that a person may feel when she makes concessions on matters of principled concern to her (Lepora 2012: 19; cf. Hall 2022: 225; Jones & O'Flynn 2012: 117; Lepore & Goodin 2013; O'Flynn & Setälä 2020: 9). In an encyclopedia entry, it is pointed out that although compromises connotate to giving up one's principles, a compromising person is able to consider the needs of other people (Menkel-Meadow 2016). A few other points have also been made, including the claim that standard moral theories do not offer a proper account of intra-personal compromises (Hoffmaster & Hooker 2017; Margalit 2010: 6). However, in general, intra-personal compromises have not received major attention in philosophy. Probably, the reason for this negligence is the view that they are much like any other demanding and regretful decision a person can face.[1] If so, there is really no need to analyze intra-personal compromises—so much has already been written about hard choices, moral dilemmas, and other psychologically demanding decision-making problems (see e.g., Gowans 1987; Sinnott-Armstrong 1988; Statman 1995).

In this paper, I deviate from the usual philosophical approach to compromises and briefly examine the notion of intra-personal compromise. As some authors have correctly pointed out, intra-personal compromises come *logically prior* to inter-personal compromises (Hall 2022: 222; Lepora 2012: 2). When a person decides to agree to compromise with someone else, she must decide which principles, values, or interests she is willing to sacrifice, partly or completely. That is, she must be ready for an intra-personal compromise, if she wants to compromise with someone else. This is why it is important to understand the nature of intra-personal compromises. Inter-personal compromises presuppose concessions. An interesting question is whether there is something special in intra-personal compromises that occur in the context of inter-personal compromises. Are the intra-personal compromises that we accept in all kinds of contexts similar to those intra-personal compromises that we make when we compromise with others? In what follows, I argue that the impression that they are similar is more or less correct. However, there are also some distinctive features in intra-personal compromises that are involved in inter-personal compromises. The intra-personal compromises that arise in the context of inter-personal compromise do not necessarily heighten responsibility but do involve extra uncertainty.

The paper proceeds as follows. I first introduce how inter-personal compromises are understood in this paper and distinguish between compromising and bargaining. Then, I present some examples of intra-personal compromises and explain what such compromises are. After that, I examine the similarities and possible differences between (a) the intra-personal compromises that we accept in all kinds of contexts and (b) intra-personal compromises that we make when we compromise with others. Finally, I draw some conclusions and estimate the meaning of the debate.

[1] "Sometimes we talk of 'compromises' in an *intra*personal fashion: Every trade-off between values can be described as a 'compromise'" (Wendt 2019: 2856).

2. Inter-Personal Compromises

When people or other agents disagree, it is often advisable to seek consensus by offering reasons that help arrive at an agreed judgment on what is right or best. However, in a non-ideal world agreed judgment is regularly out of reach, and then parties can, for instance, vote, lot, or compromise (O'Flynn & Setälä 2020: 2). If they decide to compromise, what exactly do they do?

Inter-personal compromises involve two or more parties. When the parties settle disputes through mutual concessions, they compromise.[2] If only one party makes concessions, it is no longer a compromise (Archard 2012; Day 1989). Often, compromising parties are in asymmetrical positions, which means one side is powerful, while the other is weak (Golding 1979: 15). However, such parties can compromise if both sides make genuine concessions (and do not merely pretend to give up something). If the result is strikingly unequal and largely dictated by one party, we may cease to call it a 'compromise'. However, it is not part of the meaning of 'compromise' that it is fair (Jones & O'Flynn 2012: 119). Many actual compromises are obviously unfair, although we are eager to call the results of negotiations 'compromises' especially in cases in which the results or the concessions of the parties seem equal and, in that sense, fair.

The motive behind an inter-personal compromise is that by agreeing to compromise one can gain something. The termination of a conflict is often mentioned as a goal of the compromise (Golding 1979), but parties usually try to get more than merely 'peace'. A compromise ensures that each party gets at least something they value, and making a compromise is often preferable to continuation of a disagreement. Having at least one solution is better than not having a solution at all.[3] The pragmatic justification for compromises is that it is often important to get *something* done (Menkel-Meadow 2016: 4).[4] We can say that by compromising the parties get their second-best options, as the best option is not feasible, because of the disagreeing party (O'Flynn & Setälä 2020: 9; Wendt 2019: 2856). Notice, however, that when the negotiation process is still going on, the parties do not necessarily know which options will be available—although they know that the best option is not.

We can distinguish among three types of compromises. The following taxonomy is derived from Chiara Lepora (2012: 7-9). (1) In a substitution compromise, parties agree to do something different than what either of them primarily wanted. They substitute their goals, and do not preserve even parts of the original goals. (2) In an intersection compromise, compromising parties preserve some of their principles or goals. Only a subset of the principles or goals that each held from the onset is sacrificed. In this sense, intersection compromises require fewer concessions than substitution compromises. (3) In a conjunction compromise, parties originally have completely opposing principles (or goals), but they decide to agree to compromise so that both can preserve some of their original principles

[2] "A concession is an offer by a prospective co-operator to accept a particular utility less than that of his claim". This is how David Gauthier (1986: 142) defines the term 'concession'. See also Gauthier 1990: 180.

[3] Amy Gutmann and Dennis F. Thompson (2012: 30) argue that "the greater the resistance to compromise, the greater the bias in favor of the status quo".

[4] Andrew Sabl (2018) writes about 'necessary compromises', which are compromises motivated by the knowledge that failure to meet a deadline to act would result in public harm.

(or goals). Conjunction compromises tend to be particularly unpleasant, as the parties should agree to pursue things they original opposed.[5]

Many authors have emphasized that compromises need to be voluntary—otherwise they are not compromises. We may sometimes want to say that someone was 'forced to compromise', but in these cases we (usually) mean that the compromise was especially inconvenient for some party (cf. Braybrooke 1982), or we refer to the agreements which are not literally compromises at all, such as certain peace treaties. The voluntariness of an inter-personal compromise requires that the overt act of compromising, such as hand-shaking, verbal statement, or a signature, is unforced (however unpleasant it may be). But it also requires that the parties understand what they are sacrificing and what they are probably gaining from their concessions. Notice that the decision to agree to compromise in certain way can happen much earlier that the actual inter-personal action of compromising. This personal deliberation should also be voluntary, and not based on massive manipulation or an insolent deception, for instance. Of course, compromises can and often do involve elements of dishonesty. If I gently lead you to believe that my lowest price for selling X would be Y, but in fact I would be ready to sell X to you for half-Y, that seems acceptable. It seems acceptable, as parties need not reveal their cards during the negotiations (for a discussion, see Kuflik 1979: 47).

The reason why it is important that compromises are voluntary is simply the fact that only then do they have obligatory force. To be a party of a compromise is to have freely accepted a moral commitment (Jones & O'Flynn 2020: 119). Compromises can be compared to promises. A person may promise something because that something sounds good; but if that something ceases to sound good this does not imply that the person should not keep her promise. The reason for keeping a promise is not reducible to the reason for making it (Raz 1975).[6] The same holds for inter-personal compromises. They have obligatory force even when a person is no longer willing to agree to compromise.[7] No doubt, a person may start to think that justice and fairness requires her to disengage from a compromise. If so, then she needs to decide what is morally more important: keeping the agreement to compromise or promoting justice and fairness.[8] However, in any case, an involuntary promise does not commit the promisor to keep the promise, and the same holds for compromises (cf. Owens 2006: 118).

Theodore Benditt (1979: 27) writes that a special feature of compromises is that the parties do not try to get as much as they can. However, this sounds

[5] In a substitution compromise, side A wants actions a,b,c, and d, while side B wants actions e, f, g, and h. Then they pursue actions i, j, k, and l. In an intersection compromise, side A wants actions a, b, c, and d, while side B wants actions c, d, e, and f. Then, they pursue actions c and d. In a conjunction compromise, A wants actions a, b, c, and d, while B wants actions not-a, not-b, not-c, and not-d. If they then pursue (or let others pursue) actions a, b, not-c, and not-d, they have made a conjunction compromise (see Lepora 2012: 7-9).

[6] If A wants to have a cup of coffee with B, A has a reason to meet B and may promise to meet B. Suppose that A promises to meet B. In this case, even if A would change her mind and would not want to have a cup of coffee with B, she would still have a reason to meet B, because of her promise.

[7] Successful compromises involve an element of *trust* (Golding 1979: 18).

[8] Suppose A and B make a compromise which is just and fair according to appropriate standards of distributive justice. In this case, their compromise has obligatory force not only because they have agreed to its conditions, but also because justice requires them to respect it (cf. Jones & O'Flynn 2012: 125).

unlikely. When a person needs to compromise major moral principles or human lives, she acts morally wrongly if she does not even try to make the best deal that she can. She should try to get as much as she can and if she succeeds, surely the question is still about a compromise. Of course, it is possible that the parties of a compromise do not try to get as much as they can. They can agree just because they think that, although the opponent's view is mistaken, the view makes some sense and it is therefore a *right* thing to compromise in a certain way.[9] Notice also that a person can accept a suboptimal compromise, because she has advantages from giving up a bit, such as that she feels good about the compromise, or her friends will respect her. It is not always that you try to get all you can.

In ordinary language, the results of all kinds of bargaining processes can be called 'compromises', given that the processes end up to a solution in which all parties get something, but something less than what they aimed at and was available (Braybrooke 1982: 141; Jones & O'Flynn 2012: 120). However, in this paper, 'compromising' and 'bargaining' are distinguished. Following many authors, it is assumed that inter-personal compromises are related to matters that are of principled concern to the parties, and that resolutions over minor issues (such as a dispute over the price of a used car) are typically *not* genuine compromises but rather 'deals' (Benditt 1979; Hall 2022; Lepora 2012; Lister 2007).[10] The point here is not that compromising requires 'mutual respect' while bargaining does not (see Bellamy 1999; Benditt 1979; Scott 1997). It is easy to imagine a case in which compromising parties do not respect each other—although their 'co-promise' implies that they have accepted some moral obligations towards each other. The point is that compromises are not merely *choices*, not even 'value choices', but concern matters that are considered *important*, morally or otherwise (Lepora 2012). When an employer and a trade union engage in hard bargaining over wages and working conditions and find a result that implies mutual concessions, we can say that they find a compromise (Jones & O'Flynn 2012: 120). If we do say so, we probably assume that the issue is of principled concern to the parties.

Suppose that a robber is in a position to take the victim's wallet but, after hearing some threats, agrees to take only half of the cash. Should we say that the robber and the victim found a compromise? Perhaps we should, but this is not clear (Benditt 1979: 30; Braybrooke 1982: 151). Probably, a better term can be a 'bargain', assuming that the parties did not consider the issue as principled concern. The victim would be upset, no doubt, but this does not imply that the issue was of major importance to her. She is likely to be upset simply because of the robbery and because the result of the bargaining process was so unfair. The question was of her money, after all.

3. Intra-Personal Compromises

In the abovementioned robbery example, the victim did not feel regret for the concession she decided to make. However, in genuine inter-personal compromises (that concern important matters) feelings of regret are almost defining features. This is because they involve intra-personal compromises that require painful sacrifices. A person who decides to agree to compromise with another party

[9] For a distinction between goal-oriented and principled compromises, see Jones & O'Flynn 2012: 120-121.

[10] The expression 'matters of principled concern' comes from Lepora 2012: 3.

has to solve an internal conflict of matters of principled concern. To some extent, she needs to rank her most important principles, values, and interests, as the opposing party does not allow her to pursue them all. In such cases, regret is an appropriate reaction (Hall 2022; Jones & O'Flynn 2012; Lepora & Goodin 2013). We sometimes speak about our or our interests having been 'compromised', using the very term.

In the philosophical literature, the concept of 'intra-personal compromise' has been used in two different ways. Let us call them the *literal meaning* and the *phenomenological meaning*. Those who defend the literal meaning think that intra-personal compromises resemble inter-personal compromises such that when a person's principles (or values) conflict, they should be revised and weakened so that they are compatible again. None of the conflicting principles (or values) needs to be completely rejected—just like none of the parties needs to capitulate completely in inter-personal cases. Those who defend the phenomenological meaning accept that some intra-personal compromises require people to reject certain principles (or values or interests) completely. This is often how people feel when they agree to compromise while realizing the costs of the compromise. Thus, the label is 'phenomenological'.

The literal meaning has been relatively common. Barry Hoffmaster and Cliff Hooker (2017: 55) write that an intra-personal compromise "occurs when obligations conflict and repudiating one obligation entirely to satisfy another entirely is unacceptable—for example, when a single parent cannot both raise a child satisfactorily and earn income that living together demands". Similarly, P. Anne Scott (1997: 151) argues that if an internal conflict between principles is solved by just letting one principle to override another, "this is not in keeping the meaning of compromise" and it "is not what is suggested that should happen in a compromise situation". Joseph H. Carens (1979: 125) shares the literal understanding of intra-personal compromises as follows:

> In the language of economics, one makes marginal trade-offs among available alternatives. This kind of choice is a compromise, even if purely internal one. The individual wants both the career and the relationship. He would prefer not to sacrifice any aspect of either one. Nevertheless, the need to act in a particular situation may force him to make some sacrifice, and frequently an individual chooses a smaller partial sacrifice of both of his goals in preference to a larger sacrifice of one.

The literal meaning of 'intra-personal compromise' is understandable, given what is usually meant by inter-personal compromises. Inter-personal compromises require mutual concessions; perhaps intra-personal compromises require a sort of mutual concessions as well, although the conceding 'parties' are inside a person's mind. However, the literal meaning does not do justice to our real experience of the dilemmas we sometimes need to solve. Although there are cases in which the 'partial sacrifice' of conflicting principles or values suffices, it is not uncommon that we need to put aside some important principles or values altogether. Sometimes, only these kinds of regretful major sacrifices allow us to make a compromise with someone else. Therefore, the phenomenological meaning of 'intra-personal compromise' is clearly better than the literal meaning. Unsurprisingly, the phenomenological meaning has many supporters (Hall 2022; Lepora & Goodin 2013; Wendt 2019).

Inter-personal compromises cannot be made without intra-personal compromises. But intra-personal compromises are relatively common even when they are not linked to inter-personal compromises. Let us consider some examples of intra-personal compromises that do not occur in the context of inter-personal compromises.

1. A JUDGE. Suppose that a judge has to make the final decision about whether an accused person is guilty. As a professional, she respects the presumption of innocence and thus presumes that the accused is not guilty, until the prosecutor has presented sufficient evidence in support of the view that the accused is guilty. She notices that sufficient evidence, as defined by the relevant regulations, has not been presented. However, as it happens, she is personally convinced that the accused is actually guilty (cf. Räikkä 2014: 53). This view is based on evidence, but not on evidence of the right sort. She understands perfectly well that her personal impressions that uphold her view that the accused is guilty could not be used in decision-making. Nevertheless, this understanding causes her feelings of discomfort and regret. She would like to do justice to the victims of the crime—this is one of the major reasons why she is a judge in the first place. However, she also wants to respect the institutional rules that form the basis of democratic legislation. Thus, she announces that the accused is 'not guilty'. Although the decision is psychologically painful—personally, she thinks that the accused is guilty—the resolution is not difficult in the sense of being complicated to make. In contrast, she finds it obvious to follow the guidance of the presumption of innocence. The decision she makes is related to matters of principled concern. But she is accustomed to making intra-personal compromises. She realizes that she has principles that may conflict, in this case the principle that she must respect institutional regulations and the one that victims have a right to get justice. However, she does not have contradictory beliefs. The belief that (a) the prosecutor has not presented sufficient and acceptable evidence in support of the view that the accused is guilty, and the belief that (b) the accused is guilty, are clearly compatible.[11] Some people criticize her for acting against her values. She understands the criticism, but thinks that, all things considered, she acted rightly.

2. A HOSPITAL CEO. A person in the highest-level management position in a hospital has to decide how to prioritize among different patient groups. She has a background in philosophy and bioethics and is very good at reasoning, which is required in ethical decision-making. She makes a proposal and informs about the conclusion and its grounds to relevant stakeholders: patients, nurses, medical doctors, and so on. Unfortunately, the community disagrees with her almost unanimously, and some of the comments are rather aggressive. They are not experts in bioethics and find it difficult to follow the CEO's nuanced reasoning. They prefer to stay in old practices, although they confess that sometimes the familiar policies look a bit fuzzy and arbitrary. The head of the hospital makes the following decision: old practices will not be changed. The decision is purely pragmatic, as her aim is to secure good working conditions and patient safety (cf. Devolver & Douglas 2017: 114). She does not think that it is somehow epistemically relevant that so many people have old-fashioned intuitions. She knows that ethical dilemmas cannot be solved by voting. An opinion that looks wrong in the light

[11] I would like to thank Vincent Luizzi for the helpful discussion.

of reasons does not turn right just because so many support the opinion.[12] Of course, after noticing the number of opponents, she double checks her view. However, she does not think that the disagreement as such is a reason enough to reduce her confidence in the conclusion she has drawn.[13] The decision not to reform the hospital's practices is painful, and she feels regret. She has always wanted to follow the best ethical practices that are realizable within the economic constraints. But now this principle has to go, for a while at least. She makes an intra-personal compromise and let pragmatic matters override her principle concerning the importance of ethics in hospital work. Some people blame her and say that she has sacrificed her integrity. However, this is not how she feels. Although she admits the moral costs of her decision, she thinks that, overall, she made the right decision. Hopefully, others will someday agree with her, she wonders. She is uncertain about the strength of the possible backlash in the future.

3. A CRITICAL CITIZEN. A person would like to write an opinion piece about a war. She is an experienced researcher and would have an important message. Her analysis of the situation would be original and revealing. However, as it happens, the political climate is very hostile and the media and most politicians require that (in times like these) it is important to speak with a 'single voice'. People should not present diverse opinions as such activity would endanger 'national interests'. The critical citizen considers such pro-censorship views anti-democratic (and even semi-fascists), and this gives her a further incentive to write. However, she decides to remain silent.[14] The reason for this decision is certainly not that she agrees that civic discussion is dangerous. But she understands that there is not much she can do. Presenting an analytical and historically learned opinion would not stop the war psychosis that can be seen everywhere in her country, and criticism would only harm her reputation which is presently good. It is important to keep it good, for someday she can possibly influence national policies (when times are different). The reason for her self-censorship is mainstream media and political atmosphere rather than 'extremists' or 'radicals', who are often blamed for silencing researchers. She makes an intra-personal compromise between the (a) principle that she has an obligation to correct gross public misunderstandings on important matters and the (b) principle that she should not spoil her reputation as a discerning and judicious citizen and expert for nothing. She lets the latter principle override the former, although this causes her feelings of regret. The decision is anything but easy, and it does not help that some of her friends criticize her. The criticism makes her think about whether she is just a coward and a self-deceiver.

[12] For a different view, see Earp et all 2021. The authors (Earp 2021: 108) argue that the data concerning people's actual ethical views can have 'normative implications'—and not only in the sense that, for pragmatic reasons, we should consider people's actual views. They also seem to claim that the burden of proof is on the side who opposes unanimous majority (Earp et al 2021: 99). But this cannot be right: surely the burden of proof is on the side who has not replied to the *reasons* presented in the discussion.

[13] Klemens Kappels (2018: 89) argues that "[w]hen our moral commitments are challenged by disagreement, this should sometimes make us less confident in their truth".

[14] In David Archard's (2012: 407) example a member of an advisory body should decide whether to consent to a consensus, as the weightiness of the decision "is enhanced if the committee can, given its members' differences of perspective and background, speak with a single voice".

These examples show that, at least in some cases, intra-personal compromises may require sacrifices of entire principles, not just parts of them. The above examples are rather similar but have small differences. In the 'Judge' example, the overriding principle is based on the importance of institutional rules. In the 'Hospital CEO' example, the main concern and the strongest value is pragmatic. In the 'Critical Citizen' example, the question is about omission rather than action. The examples show that deciding between principles can be but need not be easy. They also show that in intra-personal compromises a decision-maker faces non-ideal circumstances—a world that she would hope to be different—and makes the best decision that is feasible in those circumstances.[15]

4. Are Intra-Personal Compromises Special in the Context of Inter-Personal Compromises?

Let us now turn to the question of whether there is something special in intra-personal compromises that occur in the context of inter-personal compromises. Are the intra-personal compromises that we accept in all kinds of contexts similar to those intra-personal compromises that we make when compromising with others? Admittedly, they look rather similar, but let us consider three possible arguments in favor of the view that, in reality, they are slightly different.

Argument 1. When a person makes an intra-personal compromise which is also a decision to agree to an inter-personal compromise, she may realize that something morally bad can happen because of her decision—and not only because (due to the compromise) she is unable to do what is in her view the morally best option but also because by compromising she allows or enables the other side to do something bad. Compromising may thus *increase* a person's moral responsibility. In their book *On Complicity and Compromise* (2013: 26-27) Chiara Lepora and Bob Goodin correctly point out that

> a crucial part of compromise (agreeing the compromise) is a joint action, not a shared one. In a business partnership, all partners are fully liable for all the actions of all the other partners undertaken pursuant to the partnership. So too are all co-principals in a compromise morally liable for the existence of the compromise and all actions pursuant to it. […] In compromise, each party not only retains responsibility for what he himself commits and omits; he also acquires contributory responsibility for what the other does as a result of his permitting, enabling, or inducing. The greater the wrongs that you facilitate, the more blame that you share. Multiplying the numbers in that sort of case increases rather than reduces the blame that you bear.

Lepora and Goodin do not say so, but could this be a feature that distinguishes between the intra-personal compromises that we make when we compromise with others from those intra-personal compromises that we accept in all kinds of contexts? This is an interesting option, as usual intra-personal compromises

[15] In intra-personal compromises a person does not have 'contradictory' principles. The person's principles conflict merely because of circumstances, as there is no sufficient evidence of guilt ('Judge'), the community happens to disagree with her ('Hospital CEO'), and the political atmosphere is corrupted ('Critical Citizen').

increase a person's responsibility only in the sense that they tend to involve moral costs. Usual intra-personal compromises do not lead to *collective responsibility*.

However, some intra-personal compromises that are not linked to inter-personal compromises do lead to collective responsibility. Let us consider the 'Hospital CEO' example. It is not an inter-personal compromise, as it does not involve mutual concessions. However, it does increase a person's moral responsibility, as the CEO's decision makes it possible that mistaken and morally problematic prioritizing practices continue. Intra-personal compromises that are not linked to inter-personal compromises can have the same problem as the intra-personal compromises that occur in the context of inter-personal compromises. Therefore, the first argument for the view that the intra-personal compromises that we make when we compromise with others differ from the ones we accept in all kinds of contexts, is not convincing.

Argument 2. Let us consider another option. Inter-personal compromises may lead to *residual obligations*. As inter-personal compromises may include moral costs, these costs may imply an obligation to pay the costs. This point is made by Edward Hall who has especially in mind political compromises. He criticizes Lepora's and Goodin's approach and argues that "their framework presents choosing to compromise as much like any other morally difficult decision agents may make because all genuinely difficult moral decisions can be analyzed in terms of the goods and bads of commission or omission" (Hall 2022: 222). According to Hall, Lepora and Goodin's view "obscures some salient ways that political compromises differ morally from other regretful decisions ordinary agents sometimes make" (Hall 2022: 222-223). In Hall's view, compromises (at least in politics) lead almost necessarily to residual obligations, which is a special feature of political compromises. He writes that

> however a politician chooses to negotiate a weighty compromise, residual moral claims of the relevant kind are likely to arise. To the extent that representatives take their role responsibility to be faithful and committed advocates seriously they have reason not to accord with the requirements of ethics in compromise. At the same time, to the extent that representatives take the requirements of ethics in compromise seriously, they have reason not to negotiate toughly. Reasonable grievances about the way that a political compromise was negotiated can, therefore, be made from either direction even if the decision to compromise can be vindicated all-things-considered from the perspective of both parties. If a politician accords with the requirements of ethics in compromise, their representees might reasonably complain that this has come at the cost of them being a good advocate. On the other hand, if a politician prioritizes being an effective advocate, they can be charged with violating the requirements of ethics in compromise (Hall 2022: 229-230).

Could it be that the intra-personal compromises that we make when compromising with others are special because only such intra-personal compromises lead to residual obligations? This is an intriguing option, as it seems clear that the usual intra-personal compromises that we make in all kinds of contexts do *not* lead to residual obligations. If we accept Hall's assumption that the intra-personal compromises that are linked to inter-personal compromises (in politics) tend to lead

to residual obligations, then perhaps we can conclude that the intra-personal compromises that we make when we compromise with others *are* special.[16]

However, this conclusion would be too hasty. Consider the 'Critical Citizen' example. It is conceivable that when an expert decides not to contribute to the public debate although she would have an important message to tell, the decision leads to some residual obligations. At the very least, she should explain (and possibly apologize) to her friends why she did what she did. Perhaps she should also—because of her omission—do at least something. Maybe she should encourage others to participate in the public discussion. Or perhaps she should present her views to a smaller audience, or write anonymously (if it is not crucial that it is *she* who makes the point). In any case, it is possible or even probable that she does have some residual obligations due to her intra-personal compromise, although that compromise is not linked to any inter-personal compromise. Therefore, the second argument fails to establish that the intra-personal compromises that occur in the context of inter-personal compromises are special.[17]

Argument 3. As mentioned earlier, the parties of an inter-personal compromise do not know which options are actually feasible, if the negotiations are still going on. One may think that a party is not in a position to make autonomous decisions, as the other side can control which options are available. This is partly correct but notice that the problem concerns both sides. Donald Strickland (1980: 807) makes an interesting point when he argues that there is a "hidden role of coercion and fate in compromises: either third parties impose a compromise, or nature does". The role of *fate* is something that seems to define inter-personal compromises, and thus it is also present in the intra-personal compromises that necessarily occurs in the context of inter-personal compromises.

In the intra-personal compromises that are linked to inter-personal compromises there are elements of uncertainty which are absent in usual intra-personal compromises. In the 'Judge', the 'Hospital CEO', and the 'Critical Citizen' examples the decision-maker *knows* exactly what she needs to sacrifice in order to get what she wants. She knows the price of the values or principles she does not want to betray. However, this is not the case in the context of inter-personal compromises where none of the parties knows which options are available—given that the negotiations continue and no party can alone determine its direction and coerce the other party or parties. In a sense, the parties need to have trust in fate and hope for the best. The price of the compromise is known only after the negotiation process.[18] A person who is ready for an inter-personal compromise and willing to negotiate needs to act almost blindly, at least for a while—until all cards

[16] Obviously, one may want to reject Hall's (2022: 230) claim that "if a politician prioritizes being an effective advocate, they can be charged with violating the requirements of ethics in compromise". Why would the ethics of compromise demand that one is not a tough negotiator?

[17] Here I assume that residual obligations would arise however the critical citizen solves her dilemma.

[18] Of course, a person can decide what would be the highest price she is ready to pay, but this does not change the fact that the available options are not clear for her (until the negotiations are over).

are on the table. The result of the compromise, both within oneself and with others, can come as a surprise.[19]

When a person makes an inter-personal compromise, she feels and thinks in certain ways and wants and values certain things. All this is colored by a particular sort of uncertainty that seems to be a special feature of intra-personal compromise that is involved in the process. Understood in this way, the intra-personal compromises that we make when compromising with others differ slightly from the ones we accept in all kinds of contexts. The difference is not likely to be morally important, but there is a difference, at least in usual cases.[20]

5. Concluding Remarks

I have argued that the most usual philosophical questions about compromises have been those related to inter-personal compromises rather than intra-personal compromises. This paper has aimed to fill the gap in the discussion and briefly analyze the nature of intra-personal compromises. I asked whether the intra-personal compromises that we accept in all kinds of contexts are similar to those intra-personal compromises that we make when we compromise with others. I argued that they are more or less similar, although there are also some distinctive features in the intra-personal compromises that are involved in inter-personal compromises. When a person makes an intra-personal compromise in the context of an inter-personal compromise she is forced to act under uncertainty, as she cannot know beforehand what options are really available. This is a special feature. Obviously, there can also be other distinctive features. I considered only three possible arguments for the view that intra-personal compromises that we accept in all kinds of contexts differ from those intra-personal compromises that we make when compromising with others.

In the philosophical literature, the concept of 'intra-personal compromise' has been used in two different ways. Some authors think that intra-personal compromises resemble inter-personal compromises so that when a person's principles conflict, they should be revised and weakened so that they are compatible again. None of the conflicting principles (or values) needs to be completely rejected—just like in inter-personal cases none of the parties needs to capitulate completely. Others accept that some intra-personal compromises require people to reject certain principles completely. I argued that the latter is a better way to understand intra-personal compromises, as it mirrors our real-life practices in making intra-personal compromises.

Presently, in world politics and elsewhere, it is more important than ever to understand the importance of being able to compromise. Parties are more likely to reach successful compromises when they adopt a problem-solving attitude rather than a purely strategic one. They should consider their disagreement as a problem to be solved rather than a battle to be won or lost (Carens 1979: 127). The problem-solving orientation will help the parties, and intra-personal courage

[19] If a person knows of my inner conflicts, she can try to take advantage of them in settling on an agreement between us. Even if she does not, my being conflicted while the other is a bit single-minded can weaken my position.

[20] I do not deny that someone may come up with an example of an intra-personal compromise that (1) is not linked to inter-personal compromises, but (2) is still colored by uncertainty. Obviously, a person can be ignorant of available options for a variety of reasons.

should guide individual decision-makers despite the uncertainty of the circumstances.[21]

References

Archard, D. 2012. Moral Compromise. *Philosophy*, 87, 403–20.

Bellamy, R. 1999. *Liberalism and Pluralism: Toward a Politics of Compromise*. London: Routledge.

Benditt, T. 1979. Compromising Interests and Principles. *NOMOS*, 21, 26–37.

Braybrooke, D. 1982. The Possibilities of Compromise. *Ethics*, 93, 139–50.

Carens, J.H. 1979. Compromises in Politics. *NOMOS*, 21, 123–41.

Ceva, E. 2016. *Interactive Justice: A Proceduralist Approach to Value Conflict in Politics*. London: Routledge.

Day, J.P. 1989. Compromise, *Philosophy*, 64, 471–85.

Earp, B., Lewis, J., Dranseika, V., and Hannikainen, I.R. 2021. *Experimental Philosophical Bioethics and Normative Inference. Theoretical Medicine and Bioethics*, 42, 91–111.

Gauthier, D. 1986. *Morals by Agreement*. Oxford: Clarendon Press.

Gauthier, D. 1990. *Moral Dealing: Contract, Ethics, and Reason*. New York: Cornell University Press.

Golding, M. 1979. The Nature of Compromise: A Preliminary Inquiry. *NOMOS*, 21, 2–25.

Gowans, W. 1987 (ed.). *Moral Dilemmas*. New York: Oxford University Press.

Gutmann, A. and Thompson, D.F. 2012. *The Spirit of Compromise: Why Governing Demands It and Campaigning Undermines It*. Princeton: Princeton University Press.

Hall, E. 2022. Political Compromise and Dirty Hands. *The Review of Politics*, 84, 214–37.

Hoffmaster, B. and Hooker, C. 2017. The Nature of Moral Compromise: Principles, Values, and Reason. *Social Theory and Practice*, 43, 55–78.

Huxtable, R. 2014. Splitting the Difference? Principled Compromise and Assisted Dying. *Bioethics*, 28, 472–80.

Jones, P. and O'Flynn, I. 2012. Can a Compromise Be Fair? *Politics, Philosophy & Economics*, 12, 115–35.

Kappel, K. 2018. How Moral Disagreement May Ground Principled Moral Compromise. *Politics, Philosophy & Economics*, 17, 75–96.

Kuflik, A. 1979. Morality and Compromise. *NOMOS*, 21, 38–65.

Lepora, C. 2012. On Compromise and Being Compromised. *The Journal of Political Philosophy*, 20, 1–22.

Lepora, C. and Goodin, R.E. 2013. *On Complicity and Compromise*. Oxford: Oxford University Press.

[21] I would like to thank Anna Ichino, Annabelle Lever, Vincent Luizzi, Sanni Ojala, Élise Rouméas, Oskari Sivula, Susanne Uusitalo, and the anonymous referees of *Argumenta* for helpful comments.

Lister, A. 2007. Public Reason and Moral Compromise. *Canadian Journal of Philosophy*, 37, 1–34.

Luban, D. 1985. Bargaining and Compromise: Recent Work on Negotiation and Informal Justice. *Philosophy & Public Affairs*, 14, 397–416.

Margalit, A. 2010. *On Compromise and Rotten Compromises*. Princeton: Princeton University Press.

Menkel-Meadow, C. 2016. Ethics of Compromise. *In:* in A. Farazmand, ed., *Global Encyclopedia of Public Administration, Public policy, and Governance*. Cham: Springer, 1–8.

O'Flynn, I. and Setälä, M. 2022. Deliberative Disagreement and Compromise. *Critical Review of International Social and Political Philosophy*, 25, 899–919.

Owens, D. 2006. Testimony and Assertion. *Philosophical Studies*. 130, 105–29.

Räikkä, J. 2014. *Social Justice in Practice*. Heidelberg: Springer.

Raz, J. 1975. *Practical Reason and Norms*. London: Hutchinson.

Sabl, A. 2018. Necessary Compromise and Public Harm. *NOMOS*, 59, 248–79.

Scott, P.A. 1997. Compromise and Its Limits. *Nursing Ethics*, 4, 148–57.

Sinnott-Armstrong, W. 1988. *Moral Dilemmas*. Oxford: Blackwell.

Statman, D. 1995. *Moral Dilemmas*. Amsterdam: Rodopi.

Strickland, D.A. 1980. Review of Compromise in Ethics, Law, and Politics. *The American Political Science Review*, 74, 807–808.

Wendt, F. 2019. In Defense of Unfair Compromises. *Philosophical Studies*, 176, 2855–75.

It Is Impossible to Be Able to Do the Impossible

Marco Hausmann

Ludwig-Maximilians-Universität München

Abstract

Jack Spencer has recently argued that somebody might be able to do the impossible. In response, Anthony Nguyen has argued against Spencer's arguments. In this paper, I do not argue against Spencer's arguments. Instead, I argue directly against Spencer's thesis. In the first part of my paper, I develop an argument that suggests that it is *implausible* that somebody is able to do the impossible (because somebody who is able to do the impossible would be able to do something that would have incredible consequences). In the second part of my paper, I develop an argument that suggests that it is *impossible* that somebody is able to do the impossible (because somebody who is able to do the impossible would have inconsistent obligations). In the third part of my paper, I discuss and reject three objections to my arguments. I conclude that there are good reasons to believe that it is impossible that somebody is able to do the impossible.

Keywords: Metaphysics, Philosophy of action, Ability, Possibility, Counterfactual logic.

1. Introduction

Jack Spencer (2017) has recently argued that somebody might be able to do the impossible.[1] His arguments consist, to a large extent, in describing a class of cases

[1] As Spencer makes clear, when he speaks of 'being able to do the impossible' he always means 'being able to do the *metaphysically* impossible' (see e.g. 2017: 465). Spencer suggests that somebody is *in fact* able to do the impossible (not only that it is *metaphysically possible* that somebody is able to do the impossible). At the end of his paper, he goes so far to conclude that "[y]ou and I and others like us literally are able to do the impossible" (2017: 494). The aim of this paper is to argue, contrary to Spencer, that it is *metaphysically impossible* that somebody is able to do the impossible (from which it follows, of course, that *nobody is in fact* able to do the impossible).

Argumenta 11, 1 (2025): 163–173
ISSN 2465-2334

DOI 10.14275/2465-2334/20230.hau
First published: 7 November 2023

—G-cases, as he calls them—in which it is intuitively plausible to hold that somebody is able to do something impossible.[2] Here is one of Spencer's G-cases:

> *Simple G*: Suppose that determinism is true. Let h be the complete specification of the initial conditions of the universe. Let l be the complete specification of the deterministic laws of nature. Let h ∧ l be their conjunction. Suppose that G has not, does not, and will not believe that h ∧ l. G never finds herself reading a book or listening to a radio programme about the initial conditions or the laws of nature; G was home from school and sick with the flu on the day that her physics teacher covered the initial conditions and the laws of nature in class, and the physics teacher never bothered to go over the material again. We may suppose that it is fairly common knowledge in G's community that h ∧ l, that matriculating high school seniors are expected to know that h ∧ l, that many of G's classmates know that h ∧ l, and that G is one of the brightest students in her class. The proposition that h ∧ l does not exceed G's cognitive wherewithal, either in length or in complexity, and there are no special obstacles preventing G from forming the belief (Spencer 2017: 468).

It is impossible that G knows that h ∧ l. For it is, on the one hand, impossible that G knows that h ∧ l and that it is *not true* that h ∧ l (given that it is impossible to know something false) and it is, on the other hand, impossible that G knows that h ∧ l and that it is *true* that h ∧ l (given that, because of the fact that G does not know h ∧ l and because of determinism, h ∧ l *entails* that G does not know h ∧ l). Thus, it is impossible that G knows that h ∧ l. However, according to Spencer, G is *able* to know that h ∧ l. Spencer concludes that somebody might be able to do the impossible.

Here is another of Spencer's G-cases:

> I know that the actual world is actual, as do you. But not everyone knows that the actual world is actual; many have never even considered the matter. The question arises, then, whether an agent might have the unexercised ability to know that the actual world is actual. And I think so:
>
> *Actual G*: G is an actual person, a competent college student, who, as a matter of fact, will never come to believe that the actual world is actual.
>
> I think that G, in *Actual G*, like most college students, has the unexercised (and therefore necessarily unexercised) ability to know that the actual world is actual. (Spencer 2017: 479).

Again, it is impossible that G knows that the actual world is actual. For it is, on the one hand, impossible that G knows that the actual world is actual and that it is *not true* that the actual world is actual (given that it is impossible to know something false) and it is, on the other hand, impossible that G knows that the actual world is actual and that it is *true* that the actual world is actual (because the fact that the actual world is actual *entails* that G does not know that the actual world is actual). Thus, it is impossible that G knows that the actual world is actual. However,

[2] Spencer 2017: 467-69; 477-81. In Spencer's view, G-cases are necessary but not sufficient for Spencer's argument. He explains: "G-cases are only a part of my argument. The other part is a non-standard proposal about the relationship between abilities and metaphysical possibilities, and the argument gathers force only when the two parts act in tandem" (2017: 469).

according to Spencer, G is *able* to know that the actual world is actual. Therefore, according to Spencer, somebody might able to do the impossible.

Spencer's views are gaining a lot of attention in the literature.[3] In fact, Anthony Nguyen has argued against Spencer's arguments.[4] In this paper, I do not argue against Spencer's arguments. Instead, I argue directly against Spencer's thesis.[5] In the first part of my paper, I develop an argument that suggests that it is *implausible* that somebody is able to do the impossible. The general idea is that, if one's doing something is impossible, then one's doing it would have incredible consequences. It is, however, implausible that somebody is able to do something that would have incredible consequences. In the second part of my paper, I develop an argument that suggests that it is *impossible* that somebody is able to do the impossible. The general idea is that whether one ought to do something depends, among other things, on what one's doing it would amount to. However, if it is impossible that one does it, then one's doing it would amount to something incredibly good (such that it is true that one ought to do it) as well as to something incredibly bad (such that it is not true that one ought to do it). Thus, if somebody is able to do something impossible then it is true and not true that one ought to do it (which is impossible). In the third part of my paper, I discuss and reject three objections to my arguments. I conclude that there are good reasons to believe that it is impossible that somebody is able to do the impossible.

2. The Argument from Incredible Consequences

Suppose your friend is able to do something that, without violating anybody's rights, without doing harm to anybody and with everybody's consent, would turn life on earth into a paradise. What would you answer if your friend asked you whether he *ought* to do it? I guess the answer suggests itself. Of course, he ought to do it.

Suppose, however, that your friend is able to do something that would violate everybody's rights, would harm everybody and, without anybody's consent, would turn life on earth into a hell. What would you answer if your friend asked you whether he *ought* to do it? I guess the answer suggests itself. Of course not.

Let us take a step back. Suppose your friend *tells* you that he is able to do something that, without violating anybody's rights, without doing harm to anybody and with everybody's consent, would turn life on earth into a paradise. Would you believe him? I guess the answer suggests itself. Of course not.

Jack Spencer has recently argued, however, that somebody is able to do something even though it is impossible that he or she does it. In my view, Spencer's thesis is implausible. For it is not difficult to see that Spencer's thesis amounts to the incredible thesis that somebody is *in fact* able to do something that,

[3] The consequences of Spencer's views are discussed in metaphysics (see e.g. Vetter & Busse 2022: 85), epistemology (see e.g. Heylen 2020) and philosophy of action (see e.g. Hausmann 2020: 64-65).

[4] Nguyen 2020. Besides critically examining Spencer's arguments (2020: 589-601), Nguyen argues that Spencer's "thesis carries substantial theoretical costs with it" (2020: 585).

[5] In his paper, Spencer not only argues for the thesis "that an agent might be able to do what it is metaphysically impossible for her to do" (2017: 465). He also argues for the thesis "that an agent might be able to do what it is metaphysically impossible to do *tout court*" (2017: 465). At the face of it, I only argue against the former thesis. However, given that the latter thesis entails the former thesis, I also argue against the latter thesis.

without violating anybody's rights, without doing harm to anybody and with everybody's consent, would turn life on earth into a paradise.

In arguing against Spencer's thesis, I rely on very weak logical assumptions. I assume that '□ p' is an abbreviation for 'it is necessarily true that p', I assume that '◊ p' is an abbreviation for 'it is possibly true that p' and I assume that the following equivalence holds between possibility and necessity (see, for example, Garson 2013: 20):

(EQ) $\vdash \Diamond p \leftrightarrow \sim \Box \sim p$

I assume, further, that the following axioms for necessity hold (which correspond to the weakest normal modal logic K; see, for example, Garson 2013: 30):

(K) $\vdash \Box (p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$
(RN) If $\vdash p$ then $\vdash \Box p$

Note that, given these very weak logical assumptions, anything follows from an impossibility:

(IMP) $\sim \Diamond p \rightarrow \Box (p \rightarrow q)$

For suppose that it is impossible that p but *not* necessarily true that if p then q:

(1) $\sim \Diamond p$	assumption
(2) $\sim \Box (p \rightarrow q)$	assumption

If it is not necessarily true that if p then q, then it is possibly true that p:

(3) $\Diamond (p \,\&\, \sim q)$	(2), (EQ), (K), (RN)
(4) $\Diamond p$	(3), (EQ), (K), (RN)

Thus, the assumption that it is impossible that p but *not* necessarily true that if p then q leads to a contradiction:

(5) $\Diamond p \,\&\, \sim \Diamond p$	(1), (4)

Hence, anything follows from an impossibility:

(IMP) $\sim \Diamond p \rightarrow \Box (p \rightarrow q)$

Finally, I assume that 'p $\Box\!\!\rightarrow$ q' is an abbreviation for 'if it were the case that p, then it would be the case that q' and I assume that necessary implication entails counterfactual implication:[6]

(NEC) $\vdash \Box (p \rightarrow q) \rightarrow (p \,\Box\!\!\rightarrow q)$

Let us say, for the sake of readability, that something is *incredibly constructive* just in case that it turns life on earth into a paradise, without violating anybody's rights, without doing harm to anybody and with everybody's consent. In my view, Spencer's thesis that somebody is able to do the impossible is implausible because it amounts to the incredible thesis that somebody is able to do something that would be *incredibly constructive*. To see this, suppose that somebody is able to do something even though it is impossible that he or she does it:

(1) $\exists x \exists y$ (x is able to do y & $\sim \Diamond$ x does y)	assumption

As we have already seen, anything follows from an impossibility. It is, therefore, necessarily true that, if he or she does it, it is incredibly constructive:

(2) $\Box$ (x does y $\rightarrow$ y is incredibly constructive)	(1), (IMP)

[6] See, for example, Stalnaker 1968: 106 and Williamson 2007: 156. This assumption will be discussed below.

This yields the result that, if he or she did it, it *would* be incredibly constructive:

(3) x does y $\Box\!\!\rightarrow$ y is incredibly constructive (2), (NEC)

It follows that somebody is able to do something that would be incredibly constructive:

(4) $\exists x\, \exists y$ (x is able to do y & (x does y $\Box\!\!\rightarrow$ y is incredibly constructive)) (1)-(3)

Thus, Spencer's thesis amounts to the thesis that somebody is able to do something that, without violating anybody's rights, without doing harm to anybody and with everybody's consent, would turn life on earth into a paradise. Here, then, is my first argument against Spencer's thesis: It is obvious that *nobody* is able to do something that, without violating anybody's rights, without doing harm to anybody and with everybody's consent, would turn life on earth into a paradise. I conclude that Spencer's thesis is false.

3. The Argument from Inconsistent Obligations

In my view, it is not only *implausible*, it is *impossible* that somebody is able to do the impossible. To see this, recall that if somebody is able to do something that, without violating anybody's rights, without doing harm to anybody and with everybody's consent, would turn life on earth into a paradise, then he or she *ought* to do it:

(1) $\forall x\, \forall y$ ((x is able to do y & (x does y $\Box\!\!\rightarrow$ y is incredibly constructive)) $\rightarrow$ x ought to do y) premise

Recall, further, that if somebody is able to do something that, without anybody's consent, would violate everybody's rights, would harm everybody and would turn life on earth into a hell, then it is *not true* that he or she ought to do it (let us say, for the sake of readability, that something is *incredibly destructive* just in case that it violates everybody's rights, harms everybody and, without anybody's consent, turns life on earth into a hell):

(2) $\forall x\, \forall y$ ((x is able to do y & (x does y $\Box\!\!\rightarrow$ y is incredibly destructive)) $\rightarrow$ ~ x ought to do y) premise

Suppose, however, that somebody is able to do something even though it is impossible that he or she does it:

(3) $\exists x\, \exists y$ (x is able to do y & ~ $\Diamond$ x does y) assumption

By parallel reasoning as above, that person is able to do something that would be incredibly constructive:

(4) x is able to do y & (x does y $\Box\!\!\rightarrow$ y is incredibly constructive) (3), (IMP), (NEC)

Likewise, by parallel reasoning as above, that person is able to do something that would be incredibly destructive:

(5) x is able to do y & (x does y $\Box\!\!\rightarrow$ y is incredibly destructive) (3), (IMP), (NEC)

However, if somebody is able to do something that would be incredibly constructive then he or she ought to do it. And if somebody is able to do something that would be incredibly destructive then *it is not true* that he or she ought to do it. Thus, it is true and not true that he or she ought to do it:

(6) x ought to do y & ~ x ought to do y (1), (2), (4), (5)

The assumption that somebody is able to do the impossible, therefore, leads to a contradiction. Here, then, is my second argument against Spencer's thesis: If somebody is able to do the impossible, it is true and not true that he or she ought to do it. I conclude that Spencer's thesis is false.[7]

It is not difficult to see how one might extend my argument, in order to show that Spencer's thesis is not only false, but *necessarily* false. For one might have doubts about whether it is *necessarily* true that if somebody is able to do something that, without violating anybody's rights, without doing harm to anybody and with everybody's consent, would turn life on earth into a paradise, then he or she *ought* to do it. After all, it appears to be *possibly* true that life on earth *already is* a paradise. However, it nonetheless appears to be *necessarily* true that there are *at least some* (constructive) conditions such that, if somebody is able to do something that would have the consequence that these (constructive) conditions obtain, then he or she ought to do it:

(1) $\Box \exists p \forall x \forall y ((x \text{ is able to do } y \,\&\, (x \text{ does } y \,\Box\!\!\rightarrow y \text{ has the consequence that } p)) \rightarrow x \text{ ought to do } y)$ premise

Perhaps (though in my view less likely), one might also have doubts about whether it is *necessarily* true that if somebody is able to do something that would violate everybody's rights, would harm everybody and, without anybody's consent, would turn life on earth into a hell, then it is *not true* that he or she ought to do it. However, it nonetheless appears to be *necessarily* true that there are *at least some* (destructive) conditions such that, if somebody is able to do something that would have the consequence that these (destructive) conditions obtain, then it is *not true* that he or she ought to do it:

(2) $\Box \exists p \forall x \forall y ((x \text{ is able to do } y \,\&\, (x \text{ does } y \,\Box\!\!\rightarrow y \text{ has the consequence that } p)) \rightarrow \sim x \text{ ought to do } y)$ premise

Suppose, however, that it is *possible* that somebody is able to do something even though it is impossible that he or she does it:

(3) $\Diamond \exists x \exists y (x \text{ is able to do } y \,\&\, \sim \Diamond x \text{ does } y)$ assumption

It follows, by parallel reasoning as above, that it is possible that it is true and not true that he or she ought to do it:

(4) $\Diamond (x \text{ ought to do } y \,\&\, \sim x \text{ ought to do } y)$ (1), (2), (3), (EQ), (K), (IMP), (NEC)

It is, however, impossible that it is true and not true that he or she ought to do it:

(5) $\sim \Diamond (x \text{ ought to do } y \,\&\, \sim x \text{ ought to do } y)$ (EQ), (RN)

[7] It is not mandatory to interpret the argument from inconsistent obligations in terms of moral obligations. Williams (1965: 123-24), for example, distinguishes between moral and deliberative obligations. According to Williams, it is a mistake to identify "the 'ought' that occurs in […] moral judgements […] with the 'ought' that occurs in the deliberative question 'what ought I to do?' and in answers to this question given by myself and others" (1965: 123). In his view, when we know that we very much like not to do what we morally ought to do "the deliberative question can be worth asking and [we] can, moreover intelligibly arrive at a decision, or receive advice, in answer to it that is offensive to morality" (1965: 123). In my view, one might as well interpret the argument from inconsistent obligations in terms of deliberative obligations. It is not mandatory to interpret it in terms of moral obligations.

Thus, the assumption that it is *possible* that somebody is able to do the impossible leads to a contradiction. I conclude that it is not only *implausible* but also *impossible* that somebody is able to do the impossible.

This additional argument helps to make clear that one can accept my argument quite independently of one's ethical (or metaethical) views. There is no need to enter into ethical (or metaethical) debates. For as long as one grants that there are *some* constructive conditions under which it is the case that one ought to do it as well as *some* destructive conditions under which it is *not* the case that one ought to do it, one can always construct an argument parallel to the argument above by referring to these conditions (instead of the conditions specified above) and, by doing so, still get to the conclusion that it is impossible to be able to do the impossible. One can, therefore, accept my argument quite independently of one's ethical (or metaethical) views.

4. Objections and Replies

In what follows, I discuss three objections to my arguments. To begin with, one might question the premise of my second argument that it is *impossible* to have inconsistent obligations. After all, there are much discussed examples in the literature that purport to show that it is *possible* to have inconsistent obligations. Here is one such example:

> For example, I ought to protect my children from harm, and I ought not to harbor a criminal, but if my child breaks the law and I am in a position to hide him so that he escapes punishment, then it seems I ought to turn him in because he is a criminal […], and I ought not to turn him in to protect him from harm […] (Garson 2013: 48).[8]

It goes without saying that these examples are controversial (see, for example, Conee 1985). Be that as it may, this objection rests on a misunderstanding. For my assumption that it is impossible to have inconsistent obligations is not the controversial assumption that it is impossible that something is such that somebody *ought to do* and *ought not to do* it, but only the uncontroversial assumption that it is impossible that something is such that *it is true and not true* that somebody ought to do it. The latter assumption is, unlike the former assumption, uncontroversial because to reject the latter assumption means to reject the assumption that contradictions are not possibly true.[9]

[8] Garson's view is that "conflicts of obligation are possible" but that conflicts of obligation arise "because *conflicting systems* of obligation pull us in different directions" (2013: 48). For further discussion see, for example, Williams 1965, Marcus 1980 and Conee 1985.

[9] Spencer, of course, does *not* reject the assumption that contradictions are not possibly true. See, for example, Spencer 2017: 489. A qualification: Throughout my argument, I assume that one can meaningfully assign truth-values to statements such as 'x ought to do y'. Spencer appears to agree (see Spencer 2017: 487). There might, however, be non-cognitivists that would reject this assumption and that would, therefore, reject the meaningfulness of the assumption that it is impossible that something is such that *it is true and not true* that somebody ought to do it. Obviously, this is not the place to discuss non-cognitivism. All that can be said is that everybody who does not subscribe to non-cognitivism has good reasons to reject the view that somebody is able to do the impossible. Thanks to an anonymous referee for pressing me on this point.

Second, one might question the logical assumptions of my arguments. In particular, one might question the assumption that necessary implication entails counterfactual implication. After all, if *anything* is necessary implied by an impossibility then, if necessary implication entails counterfactual implication, *anything* is counterfactually implied by an impossibility. There are, however, reasons to doubt that anything is counterfactually implied by an impossibility. There are much discussed examples in the literature that purport to show that *not anything* is counterfactually implied by an impossibility. Here are two such examples:[10]

(1) If Hobbes had (secretly) squared the circle, all sick children in the mountains of South America at that time would have cared.
(2) If intuitionistic logic were the correct logic, then the law of excluded middle would still be valid.

In my view, there are two reasons why this objection is not a serious problem for my argument. First, the assumption that anything is counterfactually implied by an impossibility (as well as the assumption that necessary implication entails counterfactual implication) has not only been extensively defended,[11] it is a theorem of standard axiomatic systems for counterfactuals (such as Stalnaker's, Lewis's, Williamson's or Leitgeb's).[12] Thus, insofar as these axiomatic systems are sound and complete with respect to plausible semantics for counterfactuals,[13] this assumption appears to be justified—at least, at the face of it.

Second, slightly different arguments are available *without* the assumption that necessary implication entails counterfactual implication—arguments that still have enough intuitive appeal to make a compelling case against Spencer's thesis. To see this, suppose that your friend *tells* you that he is able to do something that has the *necessary consequence* that it turns, without violating anybody's rights, without doing harm to anybody and with everybody's consent, life on earth into a paradise. Would you believe him? I guess not. Suppose, further, that your friend is able to do something that has the *necessary consequence* that it turns, without violating anybody's rights, without doing harm to anybody and with everybody's consent, life on earth into a paradise. Ought he to do it? Of course, he ought to do it. Suppose, however, that your friend is able to do something that has the *necessary consequence* that it violates everybody's rights, harms everybody and, without anybody's consent, turns life on earth into a hell. Ought he to do it? Of course not. I take it, therefore, that slightly different arguments are available without the assumption that necessary implication entails counterfactual implication—arguments that are still intuitively plausible enough to make a strong case against Spencer's thesis.

This brings me to the third objection. One might reject the premise of my second argument that, if somebody is able to do something that would be incredibly constructive, then he or she ought to do it. One might try to argue against this assumption on the grounds that, if somebody is able to do something that would

[10] See, for example, Nolan 1997: 544, Brogaard & Salerno 2013: 642-43 and Berto, French, Priest and Ripley 2018: 696.
[11] See Lewis 1973: 24-25, Williamson 2007: 171-75 and Williamson 2018.
[12] See Stalnaker 1968: 105-106, Lewis 1973: 132, Williamson 2007: 293, Leitgeb 2012a: 54-55.
[13] See, for example, a possible worlds semantics for counterfactuals, as developed by Stalnaker (1968) or Lewis (1973), or a probabilistic semantics for counterfactuals, as developed by Leitgeb (2012a, 2012b).

be incredibly constructive, then he or she ought to do it *only if it is false that it would also be incredibly destructive*. One might argue as follows: If somebody is able to do something impossible, then it would not only be incredibly constructive, it would also be incredibly destructive (given that anything is counterfactually implied by an impossibility). It follows that, if somebody is able to do something impossible, then *it is false* that he or she ought to do it. It follows, further, that if somebody is able to do something impossible, then, *even if it would be incredibly constructive*, it is false that he or she ought to do it. There is, however, somebody who *is* able to do something impossible. There is, therefore, somebody who is able to do something that would be incredibly constructive, even though it is false that he or she ought to do it. Therefore, so the objection goes, one of the premises of my second argument is false.[14]

I guess it is obvious why this objection is problematic. This objection is problematic because it *presupposes* Spencer's thesis (it presupposes that somebody is able to do something impossible). It has no force whatsoever for those who are *not* adherents of Spencer's thesis.[15]

In my view, however, there is a further reason why this objection is problematic. For this objection has the consequence that, necessarily, if somebody is able to do something impossible, then it is false that he or she ought to do it. This, however, is at odds with Spencer's views on the matter and threatens to undermine Spencer's arguments. For Spencer's arguments consist, to a large extend, in describing G-cases. Spencer shows, however, that if G-cases are genuine cases of

[14] It is not difficult to see how one might extend this objection such as to apply to my extended second argument. For it appears to be *necessarily* true that there are *at least some* (destructive) conditions such that somebody ought to do something *only if it is false that it would have the consequence that these (destructive) conditions obtain*. It is *necessarily* true, however, that if somebody is able to do something impossible, then *it would have, as a consequence, any condition whatsoever* (given that anything is counterfactually entailed by an impossibility). It is, therefore, *necessarily* true that if somebody is able to do something impossible, then it is false that he or she ought to do it. It is, however, *possible* that somebody is able to do something impossible. It is, therefore, *possible* that somebody is able to do something that would have, as a consequence, *any* condition whatsoever, *and that it would still be false that* he or she ought to do it. It is, therefore, *possible* that *no* (constructive) condition is such that, if somebody is able to do something that would have the consequence that this (constructive) condition obtains, then he or she ought to do it. Therefore, so the objection goes, one of the premises of my extended second argument is false.

[15] To be sure, this objection is *not* problematic because it rests on the assumption that, if somebody is able to do something that would be incredibly constructive, then he or she ought to do it *only if it is false that it would also be incredibly destructive*. In fact, I do *not* reject this highly plausible assumption. Quite the contrary. I am committed to this assumption. For, in my view, this assumption is not only *compatible with* the premises of my argument, it *is entailed by* the premises of my argument. To see this, note that it is entailed by the premises of my argument that nobody is able to do the impossible and that, if somebody is able to do something that would be incredibly constructive, then he or she ought to do it. However, if nobody is able to do the impossible, then nobody is able to do something that would be incredibly constructive and that would also be incredibly destructive. It follows that if somebody is able to do something that would be incredibly constructive, then he or she ought to do it and *it is false that* it would also be incredibly destructive. Thus, it is entailed by the premises of my argument that, if somebody is able to do something that would be incredibly constructive, then he or she ought to do it *only if it is false that it would also be incredibly destructive*.

somebody who is able to do something impossible, then some G-cases are also genuine cases of somebody who is able *and ought* to do something impossible (Spencer 2017: 486-88). Thus, insofar as somebody has reasons to accept G-cases (that is, reasons to believe that G-cases are *possible*), one has also reasons to accept the thesis that it is *possible* that somebody is able *and ought* to do something impossible. This objection, however, has the consequence that it is *impossible* that somebody is able *and ought* to do something impossible. Thus, this objection contradicts Spencer's own views about the matter and, what is worse, threatens to undermine Spencer's arguments.

5. Conclusion

Jack Spencer has recently argued that somebody might be able to do the impossible. As we have seen, there are good reasons to believe that it is *implausible* that somebody is able to do the impossible (because somebody who is able to do the impossible would be able to do something that would have incredible consequences). As we have also seen, there are good reasons to believe that it is *impossible* that somebody is able to do the impossible (because somebody who is able to do the impossible would have inconsistent obligations). I conclude that there are good reasons to believe that it is impossible that somebody is able to do the impossible.[16]

References

Berto, F., French, R., Priest, G., and Ripley, D. 2018, "Williamson on Counterpossibles", *Journal of Philosophical Logic*, 47, 4, 693-713.

Brogaard, B. and Salerno, J. 2013, "Remarks on Counterpossibles", *Synthese*, 190, 639-60.

Conee, E. 1982, "Against Moral Dilemmas", *The Philosophical Review*, 91, 1, 87-97.

Garson, J. 2013. *Modal Logic for Philosophers*, *Second Edition*, Cambridge: Cambridge University Press.

Hausmann, M. 2020, "The Consequence of the Consequence Argument", *Kriterion*, 34, 4, 45-70.

Heylen, J. 2020, "Factive Knowability and the Problem of Possible Omniscience", *Philosophical Studies*, 177, 65-87.

Leitgeb, H. 2012a, "A Probabilistic Semantics for Counterfactuals, Part A", *The Review of Symbolic Logic*, 5, 1, 26-84.

Leitgeb, H. 2012b, "A Probabilistic Semantics for Counterfactuals, Part B", *The Review of Symbolic Logic*, 5, 1, 85-121.

Lewis, D. 1973. *Counterfactuals*, Oxford: Blackwell.

Marcus, R.B. 1980, "Moral Dilemmas and Consistency", *The Journal of Philosophy*, 77, 121-36.

Nolan, D. 1997, "Impossible Worlds: A Modest Approach", *Notre Dame Journal of Formal Logic*, 38, 4, 535-72.

[16] I am grateful to Kadri Vihvelin and to two anonymous referees of this journal for their helpful comments on an earlier version of this paper.

Nguyen, A. 2020, "Unable to Do the Impossible", *Mind*, 129, 514, 585-602.

Spencer, J. 2017, "Able to Do the Impossible", *Mind*, 126, 502, 465-97.

Stalnaker, R. 1968. "A Theory of Conditionals", in Rescher, N. (ed.), *Studies in Logical Theory*, *American Philosophical Quarterly*, *Monograph Series*, 2, Oxford: Blackwell, 98-112.

Vetter, B. & Busse, R. 2022, "Modal Dispositionalism and Necessary Perfect Masks", *Analysis*, 82, 1, 84-94.

Williams, B. 1965, "Ethical Consistency", *Proceedings of the Aristotelian Society*, *Supplementary Volumes*, 39, 103-24.

Williamson, T. 2007, *The Philosophy of Philosophy*, Oxford: Blackwell.

Williamson, T. 2018, "Counterpossibles", *Topoi*, 37, 3, 357-68.

Meta-Ethical Outlook on Animal Behaviours

Sanjit Chakraborty

Vellore Institute of Technology-AP University

Abstract

The nominal ground that entwines human beings and animal behaviours is unwilling to admit moral valuing as a non-human act. Just to nail it down explicitly, two clauses ramify the moral conscience of human beings as follows: a) Can non-humans be moral beings?, b) Unconscious animal behaviours go beyond any moral judgments. My approach aims to rebuff these anthropomorphic clauses by justifying animals' moral beings and animals' moral behaviours from a meta-ethical stance. A meta-ethical outlook may enable an analysis of ethical and normative views through the limit of moral motives and reasoning. Animals' sense of moral motives and their apparatus of getting involved in moral acts cannot be compared with human actions. In human moral engagement, we abide by moral paradigmatic theories and their diversified attitudes that could have conceptual and linguistic use in our societal discourses. However, animals' intentional apparatus may have the propensity to choose an act following the moral consequences (care, utility, responsibility, etc.) in their life-forms.

Keywords: Meta-ethics, Animal behaviours, Morality, Genotype-phenotype, Universal toolkit, Ecological adaption.

1. Introduction

Two primordial clauses that safeguard the moral endowment of human beings are as follows:

a) Non-humans could not be moral beings.

b) Unconscious animal behaviours are beyond any moral judgment.

This paper aims to challenge these anthropomorphic clauses by justifying animals' moral beings and animals' moral behaviours from a meta-ethical viewpoint.

2. The Crux of Moral Being: A Pattern in the Mind

The Darwinian evaluation of animals seeks to expound on different types of animals' behaviours (ethology) and characteristics that evolve in natural selection in

Argumenta 11, 1 (2025): 175–191
ISSN 2465-2334

DOI 10.14275/2465-2334/20230.cha
First published: 9 October 2023

their natural settings (ecology). There are ample numbers of animals whose physical structures, behaviours, and life-forms are different from humans. Besides, some animals whose behaviours, appearances, and collective essence of life-forms seem quite similar to humans. This brings us to an important question—'How could animals' mental behaviours adopt the evolutionary antecedent behaviours?'

As has often been pointed out, the Cartesian line-up discards the notion of any mental life-forms of animals by prioritizing self-consciousness as a unique intrinsic property of human beings. Animals are thus full of bodily events without being allied to mental processes. Cartesian dualism contains an unbridgeable gulf between human consciousness and animal mechanisms (automata). There is no psychological superiority of thought in the animal's brain (Descartes 1970: iv). The paradigmatic characteristic of thought in an animal's mind seems weaker in Descartes' writing. 'What is mental life?' To define the notion of mental life, I favour a quote from William James:

> Mental and physical events are, on all hands, admitted to present the strongest contrast in the entire field of being. The chasm which yawns between them is less easily bridged over by the mind than any interval we know (James 1950: 134).

We can find a disparity between our mental content and bodily behaviours. The Cartesian doctrine of the dualism of mind and body and its interaction looks entirely debatable since Descartes barely looks back to the question of mental content from the prospect of language. The Cartesian picture of the mind is a non-embodied substance whose state consists of thought. Besides, British empiricist Locke induces language from an external social reality that corresponds to thought. Implementing the language in the thought procedure evokes an *idea* in the agent's mind and also infuses *ideas* with reality. A fruitful communication for Locke rests on the words that correspond to *ideas*. Locke writes, "To make words serviceable to the end of communication, it is necessary (as he has said) that they excite in the hearer, exactly the same *idea* they stand for in the mind of the speaker" (Locke 1993: 270, or III 9, 6). For Locke, the concept of knowledge relies on the accumulation of experience. He believes in the initial blankness of minds. The Lockean model pins down an animal's perceptual procedure, memory system, and reason, but does not appreciate any abstraction (*idea*) process. The passive process of perception of humans and animals is sensation-based. In fact, animals have limited sensory apparatus, so their perception has a vague sense. Locke believes in the working memory of animals like a bird who can imitate a tune that they listened to a couple of days before. Even animals have the propensity to compare or discriminate between two different things in a limited sense, and we can take the example of a dog who can compare two objects like a liquid and a solid, etc. Lockean theory construes a crucial demarcation between men and animals by bringing to bear an abstraction theory that engages words to represent the *ideas*. As animals do not have access to words, they cannot construct their *ideas*. Animal's sense-centric degrees of reason may encompass intelligence and comparison, but their cognitive apparatuses are unable to grab any abstraction in general.

The rationalist account of Descartes and the empiricist account of Locke unify the identity of human beings, where rationality, consciousness, thought, and experience play a very crucial role. For Wittgenstein, the method of thinking is a conscious level of understanding of human beings. Wittgenstein clari-

fies, "We say only of a human being, and what is like one, that it thinks" (Wittgenstein 1953: 360). So, the attribution of thought to animals depends on the representative criterion of human thought. Here, the interesting point is that we could not judge animals' mental lives as a stereotype inference or induction. The process of philosophizing the nature of intention, cognition, and memory intends to set up our mental contents from the narrow (intrinsic) and broad (extrinsic) levels. Later Norman Malcolm resonates that these self-absorbed and intrinsic procedures are indeed tangible (physical, decision-centric, sensation, and images) and intangible mental phenomena *constructed by tangible one*s (Malcolm 1976: 36-38). If we put the jargon into the Lockean model, it looks appealing because Locke believes that *ideas* come first, and names come afterwards. This hypothesis confronts a well-known hypothesis that 'language precedes thought'. The notion of meaningfulness of a term undermines the denotation part as this relies on the connotation part that animals barely advocated.

The metaphysical understanding of agency and its psychological speculation on mental contents extensively oversimplified the notion of animal thoughts. The anthropocentric outlooks against animals' thoughts and their mental lives are considered as non-thinking beings. Taking up the point, we can question their moral agency from a behaviouristic-cum-ethical interpretation. Here the ground objections are as follows:

- Animals cannot be the first-person authority or morally responsible for their behaviours and thoughts.
- They are not conscious of their actions.
- Animals could not have any mental states.
- Animals are non-linguistic, agency-less, and mindless living bodies.
- Animals cannot expose valuing, the concept of the capabilities approach, equality, and justice, so there is no question about their moral conscience.

Before tracing back to these objections, let us point up the pattern of mind and the pattern of moral agency differently to revisit the objections in defence of the justificatory clues of animals' moral behaviours and their thought-centric life-form.

To understand the patterns of the mind, language, and mental grammar, one has to review the whole scheme. Mental grammar tells us how we can put words together into a sentence to facilitate the language-communication pattern that can be socially acceptable. Mental grammar is not deliberately accessible to children. Brains genetically run our mental grammar. We cannot teach our children mental grammar from the outside; they can only interact with the sentences just by hearing them. The process of understanding and the construction of new sentences indubitably depend on the pattern of the mind, which may be called innate-based mental grammar. I prefer to bring in the notion of *genotype*, an inherent biological mechanism that conduits frequent changes in humans' biological adaptation. The biological evaluation links with *genotype* rather than *phenotype*, or certain traits. The behaviourist explanation underrates phenotype fitness, but values *genotype*, which looks at the "biological basis of all behaviour" (Wilson 1975: 4). This thesis of Wilson calls for there being a genetic constituent to the behaviours of all species. The rudimentary part of the thesis stresses the perpendicular outline that insists on a journey from genes to behaviour. Besides, the horizontal account of the thesis uplifts an ability to comprehend other minds that looks like a good exposition of human behaviours based on their socio-historical engagements. No tie-up may be accessed

between the *genotype* and the *phenotype* traits. This slight gap raises the possibility of drastically inflating the gene and deflating the mind. Flanagan clarifies:

> The sociobiologist's opening gambit is largely a rhetorical flash-in-the-pan. Ironically, it fails to increase the plausibility of tight genotype-phenotype fit precisely because it tries to elbow out of the way the most likely candidate for creating a genotype-phenotype gap (Flanagan 1984: 258).

In closing the *genotype-phenotype* gap, Wilson intends to secure the "grail of a unifying theory of biology and the social sciences" (Lumsden and Wilson 1981: ix). To articulate the mind-centric social behaviours of humans, we can expose the genes as rule-makers and the mind as rule-followers by synchronizing their functions. The *genotype*-based rules go towards prejudice since the agent has its adaptive choices (mind-based). The collective choices of the different agents and their cognitively stimulated behaviours generate a socio-cultural sphere. The concept of epigenetic rules and their expanded variations appear as the dissent of minds and human behaviours. It looks promising that before language learning, a child first learns various social interactions and expressions from elders. Children are first acquainted with the social and the conceptual world and later map out the linguistic expression and its diversified connotative and denotative parts. The method of connotation and denotation in its preliminary form concentrates on its verbal expression, and later, it aims to fix its precise reference in communication by denoting the proper referents.

To understand animal minds, one must look at the environments where animals evolved. The evolutionary theory inscribes that animals are biologically equipped with a type of mental tool to solve environmental problems or other necessities. There is a tool for animals that we may call a 'universal toolkit', which eventually helps the animals be acquainted with objects, navigate, etc. Marc Hauser says:

> Divergence from the universal toolkit occurs when species confront unique ecological and social problems. Thus, for example, bats echolocate using a high-frequency biosonar signal, but we don't. Unlike humans, bats confront the problem of navigation in the dark. As a result, they evolved a brain that is specially designed to process high-frequency sounds [...] The only way to understand how and what animals think is to evaluate their behaviour in light of both universal and specialized toolkits, mechanisms of the mind designed to solve problems. And the only way to evaluate the validity of this approach is to test our intuitions about animal minds with systematic observations and well-controlled experiments (Hauser 2000: xv).

Here the germane query is how morality could evolve in our socio-cultural paradigm.

Like our scientific ability or a mathematical equation, morality evolves because of a genetic fitness stratagem. Our moral simpliciter has an innate sensor that unconsciously impinges on our conscience. I will revert to this issue again.

3. Individual and Valuing

Before engaging in the query of whether animals can execute moral actions (or whether animal behaviours hold any moral conscience), let us clarify whom we

can call an individual and how individuals enrol themselves in the sphere of moral values. An individual is a single numerical identity of a conscious being who passes through different qualitative aspects but still holds the rationality and mindset to carry out the first-person authority of their action and life-form. When we care about the rational or mental life of an individual's performance, the debate emerges in a different circle that we may call a moral agency.

Moral principles in Kant's outlook are tied to an *ought to do* hypothesis instead of an *is to do* hypothesis since moral laws are more prescriptive or evaluative than the mode of description. The groundwork of these moral laws is beyond the sensory experience but exhibited on the pure reason that Kant calls *a priori*. Instilling moral laws from an experience-centric level cannot establish what 'one ought to do' from a universalistic sense. One must allow human's free-will and the corresponding experiences that will stimulate them to do specific moral actions preserving the categorical imperative. In the usage which Kant prefers,

a. Moral requirements are rational requirements.
b. Rational requirements go towards universal laws.
c. These moral requirements must be followed by equality which Kant calls the *categorical imperative*. Only a universal principle could be applied to all rational human beings in an equal sense (Chakraborty & Misra 2022).

Kant expounds upon the supreme principle of *practical reason*, which he calls the *categorical imperative*. This principle can be expressed in a few different formulations, the first of which runs as "Act only accordance to that maxim by which you can at the same time will that it should become a universal law" (Kant 1988: 4:421). The principle we have just quoted, then, means that people should simply promote as rules of living for themselves about those rules that everyone could always follow.

We need to notice that to become a self-conscious moral being, an individual should be responsible for their actions and duties. Moreover, an individual should be concerned about their moral rights and gives value to other human beings and non-human beings' rights as well. Although animals have this type of individuality, they do not endorse any consent to express their concerns or rational choices. Animals are not individuals who have free-will and the notion of categorical imperatives in their life-forms to understand justice and moral systems. The ideas of justice, responsibility, blame, crime, sympathy, and virtue are human-centric qualities that an animal cannot accomplish in nature. The general approach tells us that we can pet or train the animals without their consent because they are not self-conscious or aware of their states of mind. Animals are less reasoning beings who are not cautious about their authority, self-hood, and moral rights.

In fact, individuality comes through a history of our mental and physical journey. The history of our mental life has been secured by rationality, reasoning, and causal efficacy with the society where the individual belongs. The idea of individual agency consists of interaction with others. An individual is living in the field of other individuals. A rational explanation of individuality arises when individuals primarily understand their existence in a meaningful way and become responsible for their autonomous actions. Animals cannot face any moral dilemmas since they are outside of any moral sagacity.

The emerging question that crops up is how the manifestation of individuality or personhood sounds related to the attribution of subjectivity that only belongs to human beings, not in any way to animals. The divination of human behaviours is that we can only ascribe the attributes to others as a premise that we have experienced ourselves. In Strawson's work *Individuals* (Strawson 1977), he attempts to jot down a person as an Individual concept by discarding the dualistic form like the person is the amalgam of two separate entities—physical character and consciousness. The identity theory that Strawson secured describes the inseparable characteristics of a person (consciousness subject) with their attributive physical properties. Strawson believes that individual consciousness could exist as a secondary concept (non-primitive) since its existence relies on the individual or the person. The process of ascription also intertwines with other persons. Strawson argues:

> […] that it is a necessary condition of one's ascribing states of consciousness, experiences, to oneself, in the way that one does, that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself (Strawson 1977: 99).

But this analogy seems opposed to the argument of attributing experience to others by identifying their conscious patterns since an agent cannot attribute the experience to others, considering them as pure consciousness or a collection of experiences. To predict consciousness in the direction of ego subsequently envisages the state of mind collected to the subject's material body. These types of arguments shed doubts on the justification of allowing for the existence of *other minds*. Having said this, we must, however, appreciate the advantage of Strawson's argument. Strawson describes two types of predicates—*M-predicates* attributed to the physical characteristics and *P-predicates* ascribed to states of consciousness. All *P-predicates* are not ascribing the state of consciousness fully. Although these *P-predicates* entail the ownership of consciousness on the part of that to which they are attributed. *P-predicates* look, in essence, both self-ascribable and other-ascribable. One can ascribe *P-predicates* to others on the potency of their behavioural patterns, but to oneself does not depend on the behavioural criteria. Strawson believes in the criterion of any particular psychological state or property that is identical to the criteria of applying physical properties unless the agent has already been identified as a person. Strawson writes:

> But it is essential to the character of these predicates that have both first and third-person ascriptive uses, that they are both self-ascribable otherwise than based on observation of the behaviour of the subject of them and other-ascribable on the basis of behaviour-criteria. To learn their use is to learn both aspects of their use (Strawson 1977: 108).

From a critical perspective, this learning procedure is undoubtedly a synthesis of thought and language that no animals can ever attain. The concept of other minds and the attribution of other-ascribable criteria based on behavioural criteria are not available to animals, as they may have sensation or introspection, but they do not have any logical induction.

It seems, indeed, a happy outcome to consider that the aspect of valuing others as much as valuing own-self narrates a way to individualistic ownership of self-valuing of the humans and the universalizability predilection of values (a

type of value-conferring) that one may likewise engender to others. Animals are, in principle, inept at inferring values or logical induction.

4. Animal Behaviours and Morality

Descartes' famous argument against animal behaviours asserts that animals cannot master human language, so their behaviours are nothing but mechanical (automata). The literature (Hunter 1913) of the last century on animal behaviours cued stimulus representation during behaviours. Hence, for the animals, representation is there, but there is no availability of internal stimulus. The logic of parsimony illustrates that the predication or the ability to explicate behaviours does not rely on the stimulus representation, but the stimulus remains available when the individual replies to these. The fundamental difference between animals' and humans' representation looks inadequate while we underscore human-centric appeals. The recent literature shows that the initial studies on human memories are nothing but verbal stimuli. The overlapping of the encoding process between humans and animals causes the verbal rehearses of stimulus. Here the cognitive processes expand to encode the memory that looks conical for animals instead of humans. When we probe into the biological aspects, one thing is strongly clear that animals are not humans as they are different species.

If we recapitulate the Aristotelian meta-ethical claim, in a nutshell, virtue is grounded in human endeavour. It is thus a practice exclusively accomplished by human beings.[1] Spinoza argues:

> The principle of seeking our advantage teaches us to be in close relationship with men, not with beasts or things whose nature is different from human nature, and that we have the same right over them as they over us. Indeed, since every individual's right is defined by its virtue or power, man's right over beasts is far greater than their right over men (Spinoza 1982: 4p37sl).

As human moral evaluation and conviction are controlled by reasoning and intentionality, we cannot compare these with animal behaviours. We cannot blame animals for their actions, while we can blame humans for their reason-based acts. Animals are beyond any moral deplore or moral admiration as their behaviours outride any *self-imposed restraint* or psychological and psychiatric assurance.

Human moral behaviours and attitudes endure logical and rational thoughts, whereas animals' mind-sets and behavioural dispositions are imprecise because they cannot be the agents of their life-forms. The ethical indemnity is just for intelligent animals rather than stupid and alien ones. This anthropocentric stance vividly played its own dice in the history of the human-animal interface. The concepts of reasoning, rationality, and moral conjectures, all these criteria from the Cartesian level to Kantian analyses remain challenging to the

[1] Nussbaum writes: "Aristotle's scientific spirit is not the whole of what the capabilities approach embodies. The approach includes, in addition, an ethical concern that the functions of life not be impeded, that the dignity of living organisms not be violated. Unlike Greek thinkers in the Platonist tradition, Aristotle seems not to have pursued such thoughts" (Nussbaum 2007: 348).

animals, but suffering is an explicit criterion where animals and humans merge as living beings. According to Jeremy Bentham, the justification of suffering is the criterion that uplifts animals considering the ethical domain. Many philosophers urge that the limitation of animals' thoughts is viable just because of their inability to language acquisition. However, it looks right that on account of their universal toolkit, animals can execute specific actions like recognizing objects, finding out the shortest route to reach shelter, or determining the number of individuals in a group, etc., which look comparable to human beings. The promising point is that the animals have three types of exploring mechanisms that explicate their behavioural approaches along with their adaption to nature. These mechanisms are as follows:

a. Universal toolkit
b. The mental toolkit that relates to their psychological cognition
c. Ecological adaptation.

The evidence of animals' suffering involves three different states—behaviour, cognition, and health (Dawkins 1980, 1985). Like humans, they have direct access to their cognition and feelings. We may access indirect evidence as a third-person authority of their behaviour and thought. The foremost sign of suffering comes in times of crisis of health like diseases, injury, and emaciated, but any physical damage can be suppressed in some cases because of the high capability of tolerance even in the case of humans like animals. However, evidence of the severe pain of the animals and humans can be expressed through their physical and mental aptitudes. Any physical or psychological harm makes a subsequent change in the level of hormones, brain activity, or heart rate of the animals that any physiological measurement can well measure. The interaction between the mind and the body problem becomes startling due to the supervenience theory, where an external sensation represents cognition in the subject's mind. Any physical damage or injury because of the spinning of neurons reciprocates a sensation to the brain of the animals or humans, and this sensational interaction of the mind-body instils a change in their psychological aptitudes (emotions and feelings)[2]. Moreover, behaviours look for the best evidence to understand this psychological or cognitive apparatus of animals. From Darwin's *The Expression of the Emotions in Man and Animals* (Darwin 1965) to some recent thinkers (Hauser, Dawkins) believe in the behaviour-centric account of animal's actions. They believe in animals' feelings, and how animals express their feelings in their actions articulates that animals have their minds. Animals' feelings about pain

[2] Seeing other animals in distress or painful situation, some animals feel discomfort or sympathy which is a type of moral emotion. These emotion-centric animal behaviours fulfil the significant criterion of the moral agency once wonderfully put forward by Rowlands as "X is a *moral subject* if X possesses (1) a sensitivity to the good- or bad-making features of situations, where (2) this sensitivity can be normatively assessed, and (3) is grounded in the operations of a reliable mechanism (a 'moral module')" (Rowlands 2012: 230). My understanding points out animals' sensitivity to other animals' distress and psychological suffering from the prospect of moral internalism that holds rationality and moral reasoning eventually to deploy intention-based moral choices, while the same sensitivity of the animals renders an account for moral externalism by bringing the notion of external factors (good for all) and collective agencies.

and suffering always confer them an experience to avoid unpleasant situations and live with a capacity to do something for their survival value.

What is moral right? Any kind of right is considered against incursion so that the interest that has been given to a person or agent can be secured. People have to secure their lives and interests. Because of this reason, they work out in society and bring up moral decorum. Scruton defines moral beings by saying that

> The moral being is not merely the rule-governed person who plays the game of rights and duties but a creature of extended sympathies, motivated by love, admiration, shame, and a host of other social emotions (Scruton 2000: 33).

5. Foul and Goal

In his paper "Thought and Language" (Putnam 2022) Putnam claims that animals' thoughts or understanding of objects have no fundamental concepts or thoughts like human beings. Animals have *proto-concepts* and *proto-thoughts* that only can be successful or unsuccessful. Here the point is that for Putnam, the proto-concepts of animals' thoughts can merely be satisfied or dissatisfied due to the non-relation between the claims of justification. Human reason-centric thought consists of truth values that could be justified and unjustified.

Let us elucidate what thought is and how thought can be distinguished from proto-thoughts. Thoughts are the reason-oriented mental contents of human minds that help an agent to understand truth values from the level of justification. The pre-linguistic animals have some precision settings that aid them in recognizing the Pavlovian reflexes but do not hold any true mental representation. Putnam argues that as pro-concepts look indeterminate, we cannot question their justification or truth values. Putnam feels uneasy about inculcating any possible relation between pro-concepts and innate language. Determination or any reason-based justification is the basic sign of concepts. The interesting point is that if we compare human beings' visual systems with animals, we will realize that in their case (animals) the visual systems and the colour perception processes are not the same as humans. That's why it would be fallible to argue that when an agent perceives a crocodile and at the same moment when the crocodile also perceives the agent, here they have applied comparable visual processes and conceptual representation. Human visual organisms and how they perceive an object cannot be parallel to the reptile's perception. Human perception has some conceptual representations that reptiles do not have. Putnam considers:

> In *Renewing Philosophy*, on the basis of my thought experiment with the dog who is given a veggie steak, I concluded that mental representations that can be *correct* and not merely *useful* do presuppose language. If that were the whole story, then, since thought properly so-called must be truth-apt, the answer to Chakraborty's question would be that thought presupposes language *period*. Thought does presuppose language (or something very much like language, e.g. "mentalese"), *if* one requires that "thoughts" consist of propositions, can be joined with logical connectives, are components of inferences, etc. [...] But (I learned from Burge) it is not the case that only mental states that consist of propositions, or that contain concepts, can be accurate or inaccurate. *Perceptual representations* can be accurate or inaccurate. If "thought" means conceptual thought and "language" refers to

> an activity that uses concepts, then "thought presupposes (something like) language" is a truism (Putnam 2022: 54-55).

However, I think, reasonably that some animals have thought procedures, and they can think accordingly. Still, I am afraid to say that animals have the metacognitive privilege or the cognitive ability to think about their thoughts.

It seems true that conscious beings have a sensational state, but it does not mean that sensational states (pain, pleasure, etc.) can always be deciphered into propositional content. Benthamian utilitarianism taught us about preserving animals' utilities as animals can feel pain-pleasure and suffer accordingly. Human mental states insist on propositional contents, and the processor of the mental content presents the content of the belief about something that we may call the referent of the particular belief. For me, contents are in nature the referential constituent of beliefs. Besides, the causal referential directness to reality is considered a linguistic unit. However, the constituent of the linguistic entities is unavailable in animals' thoughts. Animals' thoughts do not appear to carry any referential linguistic content. Related to this conception Chomsky writes:

> The examples of animal communication that have been examined to date do share many of the properties of human gestural systems, and it might be reasonable to explore the possibility of direct connection in this case. But human language, it appears, is based on entirely different principles. This, I think, is an important point, often overlooked by those who approach human language as a natural, biological phenomenon; in particular, it seems rather pointless, for these reasons, to speculate about the evolution of human language from simpler systems—perhaps as absurd as it would be to speculate about the "evolution" of atoms from clouds of elementary particles (Chomsky 2006: 61).

The belief procedure attains a holistic purview, where beliefs can be understood as forming corporate bodies, i.e., they are causally allied to each other. Besides, the believer has an individual agency where the linguistic agent carries out agency and takes responsibility for their belief-centric actions. We can exclude here the marginal cases like autistic personalities, children, and mentally challenged persons. From a nominal account, this approach seems closer to human subjectivity, a trust in humanism. Sartre says: "And at the point of departure there cannot be any other truth than this, I think, therefore I am, which is the absolute truth of consciousness as it attains to itself" (Sartre 1963: 44). The universe that Sartre depicts in his literature represents the human universe to which the animal minds have no access. Sartre eliminates the idea of basing human thought on *animalitas*. The world and human minds reframe an intersegment of the world disclosing to the human consciousness, and this deformation is possible because of language use. Even the idea of world understanding remains explicit because of language acquisition. For humanists, the world is doubtlessly human, and humans are the only beings who can understand the 'world as world'; no other animals have this sort of comprehension. We can remember the celebrated dictum of Wittgenstein, 'I am my world'. This solipsistic cum humanistic doctrine collides with coinciding an animality into the sphere of an idealized ontology of being and its correlation to others (beings). The world that animals uphold seems to be a non-conceptual world. Just having access to sensory experience, animals' stimuli cope with the entities of the world without

knowing what the entities really are. An animal feels hungry like a human, but it cannot feel poverty, justice and pride as we often do. I think this radical anthropocentric account has foreclosed any comprehensive assessment of the organic and biological similarities between humans and animals.

If we try to see 'reason' from two diversified senses—theoretical and practical, according to Blackburn the theoretical reason is comparable to a 'navigator's map' (Blackburn 2009: 53) that could help an agent to adjust the belief system and the action afterwards. In comparison, practical reason motivates the agents to select the action, not to construct beliefs. The inevitable patterns of animals' thoughts may have a practical reason that instates them to do some actions or choose a particular thing between different alternatives, but they cannot react by the mode of theoretical reasoning like the knowledge of *apriority* (independent of experience). Animals do not envisage the *a priori* knowledge of how a thing happens to stand without experience. But one can argue in defence of animal minds that the conventional aspect of the rules of any game can be changeable because of the situation and needs, similarly, animals without engaging in *a priori* knowledge just due to their experience and inference could make out the practical reason-based actions. Many people in the world execute these types of actions without being related to *a priori* thought. Alternatively, in a Kantian sense, the notion of practical reason holds an argumentative ground, where the agent preserves certain rational deliberation, a deliberation that generates free-thinking, while rationality initiates actions based on reason and rationality. Animals' choice-based actions that may be guided by reason in a practical sense are unchained from the bond of arbitrary, irrational acts. Only humans who are privileged under moral laws and reason may produce the will that is good in itself. Humans do not always follow moral laws, but they have the integral *a priori* aptitude to follow reason-based behaviours. Humans have an autonomous impulsion that initiates us to act autonomously guided by goodwill and freedom, but the impulsion that instigates animals for their preferred actions is derived from their heteronymous unguided reasoning (if any). However, I do object to this understanding of animals' thoughts and actions since it perversely looks human-centred. The idea of believing something and practising these same things accordingly does not mean that one has to be a linguistic animal. Animals have their practical cognition of something, and they can act on the same action repetitively and could change the same action if the situation disfavours them. We have examples beyond their limited access to thinking and guided evolutionary instinct animals (especially dogs, and chimpanzees) in performing certain actions through their compassion, feelings, and previous experience. I find myself here pushed to the new domain. Animals' essence needs to be verified by more eloquent theories since the old anthropocentric modules of consciousness, linguistic being, or thinking are looked unappealing in the case of initiating a dichotomy between humans and animals. To explain adequately, animals can be the 'subject of a life' in Regan's sense, but it seems quip to call them the 'subject of morality' as they are not aware of their belief model and individual moral agency. Here, I want to defy Regan's stance. Although I never surveyed it or did not have any concrete evidence supporting me, I intend to advocate moral personhood to animals from a philosophical milieu. Finding out some routes to confer the concept of person to the non-humans, I was motivated by the thought of Rowlands, who once argued in defence of animals' personhood and wrote:

> To qualify as a person, in the eyes of this book at least, is to satisfy the four most commonly cited conditions of personhood: conditions that can be plausibly regarded as individually necessary and collectively sufficient for an individual to qualify as a person. Many animals qualify as persons by virtue of their being conscious, cognitive, self-aware, and other-aware individuals: they are individuals in which these four features coalesce. Because they are such individuals, there is no respectable reason to think of them as anything other than unified subjects of often quite complex mental lives—aka, persons (Rowlands 2019: 195).

One puts a question: can animals refer to a particular subject or object? As we know, the reference looks as if at a relation between the agent's mental states and the objective world. To get a better reference for a refereed object, one must plunge a certain propositional attitude that seems linguistic or indexical. If we delimit the access of mental representation from the tune of language, the intentionality and the constructive part of the representation preview a teleological account, where far-seeing the mechanism initiates a cognitive status from the aspect of the relational proper nouns derived from the environment. The purport of being close to environmental accessibility pins down the synchronizing status of a biological mechanism that produces the relational proper functions. Rowlands aptly argues:

> Representation, therefore, is ultimately a biological notion. And, given that non-human animals clearly have internal mechanisms that have evolved to detect certain environmental features, it is perfectly appropriate, at least in principle, to make transparent attributions of beliefs and other propositional attitudes to them [...] We know, on evolutionary grounds, that dogs are going to have evolved mechanisms to detect friends from foes, familiar from stranger animals, pick members from outsiders (Rowlands 2009: 217).

Animal's moral agency is a theoretical prospect engaged with some empirical plausibility as argued by Korsgaard (2006) and Rowlands (2017).

The primordial questions of our concern are what makes us primarily different from the animals and what construes animals not to carry out moral agency. An unnerving way persistently reveals that we (humans) are set apart from animals regardless of bodily kinship because of an abyss concept that goes towards subjectivity and nothingness together. The origin of forming the conceptual analysis persistently disdains the incommensurable opacity of the animal's moral responsibility. Humans have maintained a bit of inhumanity through their conceptual forming and behavioural patterns, but animals' ethical innovation cannot tussle with this problem. No ethical innovation can obsess the gulf between morality and non-humanity. Human prevails over animals' nature by manifesting linguistic behaviours and moral improvements.

In defining moral autonomy, the philosophical discourse spells the considerable amendment of individuality, moral choices, and responsibility. This individual agency transmits moral autonomy to others and holds a responsible sense for their own moral choices and actions. Comprehending the free choice from external constraints, Hume anticipates that an agent cannot be the final authority of their acts, but they have the freedom to act without being motivated by external constraints. However, the Kantian idea of freedom deviates from external constraints as the theory also deduces from the transcendental apparatus. It looks true to Kant that an agent can act by choosing the moral laws autono-

mously, and here the rudimentary point is that the moral laws need to be freely preferred. The coextensive of autonomy and negative freedom (freedom overseen by external forces) cannot supervene each other. Moral theorists prioritize dispositional autonomy, a necessary condition of autonomy where the second-order reflection of the agent's act supervenes on the first-order preference, which has been fully chosen by the agents themselves. The limit of the dispositional autonomy incumbents to the matter of degrees that could be gradually developed like in the youths' case as they gradually developed their first-order preference concerning the second-order reflection. How old they are getting, their preferences, and the reasons for selecting a particular preference could be gradually enhanced. But this theory is not free from moral dilemmas. Sometimes in a particular framework, the agent's first-order preference that preserved their authentic autonomy could be turned out as wrong and miserable for their future dispositional autonomy like the examples of some religious practitioners who later engaged people in fanatic terrorism in the name of blind religious faith. In the evolution of moral acts, autonomy can coexist with free-will. A recurrent confusion again is generated here: Although animals have their preferences, do they uphold any autonomous moral agency? A possible solution is to claim that animals have a first-order preference for their choices, and subsequently, they may have a second-order reflection in their mind as a prerequisite of their autonomous choices. However, it seems doubtful that animals preserve any moral agency as a dispositional autonomy. In defining moral autonomy, the philosophical outline stresses the considerable amendment of individuality, moral choices, and responsibility. This individual agency transmits moral autonomy to others and holds a responsible sense for their own moral choices.

As noted above, a person's moral claim holds a propositional attitude while the animal's mental claim has a sensational attitude. The pattern of the propositional attitude deciphers the dispositional behaviours that are interlinked to the consciousness and mental content of the person, but animals' sensational states tie to consciousness without bringing the state of mental contents. In this connection, a question arises: Can concepts stand for the context-free representation of minds? Animals can perceive the concepts, but the point seems intriguing when we intensify the range of perceptual entanglement to the mental representation. The element of our mental content corresponds to the perceptual representation. If we pursue the model with human design, we run into other problems. The facet of our mental content and the perceptual replica conjointly insert the context-sensitivity, self-knowledge, and individuality. Those tripartite codes of thought perhaps never could be attained by animals.

Meanwhile, we should note that the animals (chimpanzees and bonobos) can foresee other chimpanzees' or bonobos' intentions and actions. Still, they have certain compatible socio-cognitive abilities that are close to the phylogenetic base. Chimpanzees adopt mutual understanding and collectiveness when they are hunting or feeling benefits in teamwork.[3] Here animals' socio-cognitive skills

[3] Rowlands writes: "Jane Goodall observed chimpanzees in the Gombe Stream National Park using grasses and twigs to fish termites out of termite mounds. The chimps would often modify the twigs, by stripping off their leaves, so they could be used in this way—demonstrating, like the New Caledonian crow Betty, a facility for tool construction as well as tool use. Not far away, and around the same time, Kinji Imanishi's team found chimpanzees using rocks to crack nuts" (Rowlands 2019: 87). Using different tools is a

are much more engendered to the cooperative function instead of the competitive one. Grueneisen, Duguid, Saur, and Tomasello write:

> Importantly, the current findings provide some first strong experimental evidence that chimpanzees and bonobos are able to use their social cognitive abilities in order to successfully coordinate decisions with conspecifics in a cooperative context [...] This corresponds to the recent finding that wild chimpanzee selectively informs ignorant group members of danger which also points to the conclusion that chimpanzees can recruit their understanding of other's mental states more flexibly and across social contexts (Grueneisen, Duguid, Saur, and Tomasello 2017: 6).

Besides, any position that qualifies as conceptualism delineates valuing in a normative framework and expands the boundary of the linguistic circle closer to starkly human-centred. According to this view, the representational appeal of conceptualism exposes the second nature of human beings once we bring up the conception of moral values. This conceptualist account of linguistic nuance enshrines a reason-based rational ground representing the valuable laden theory in human moral acts. Crary urges:

> To say that animals of different kinds are "inside ethics" in my sense is to say that, insofar as they are observable, they are endowed with moral qualities. This is not the same as claiming that things animals do invite moral assessment. Nevertheless, I want to close this chapter by underlining the fact that my argument makes room for the possibility that the actions of some non-human animals are open to such an assessment. The upshot of the argument is that a conceptualist outlook does not commit us to hold that animals of all kinds are either mere bundles of stimulus-response mechanisms or mere systems of exploitable instincts. Animals may traffic in concepts, thereby occupying partial stages of rational development. One reasonable way to gloss this claim is to say that animals may be distinguished by partial forms of freedom in virtue of which their actions are rightly subject to some types of moral evaluation (Crary 2016: 120).

A danger that promotes a radical view in philosophy by amplifying the belt of thought and understanding is glomming on morality from a humanistic demand. The psychological discourse of our venture on ethics demarcates non-human beings or animals as non-rational cum non-linguistic beings who cannot entail or understand the fact or the true consequence of values. The impetus comes from Thomas Nagel's writing which intends to exemplify the motivational theoretical pattern of ethics by undermining the pattern of biology. Nagel admits:

> Ethics, though more primitive, is similar. It is the result of a human capacity to subject innate or conditioned pre-reflective motivational and behavioural patterns to criticism and revision and to create a new form of conduct. The capacity to do this presumably has some biological foundation, even if it is only a side-

part of causal reasoning that some animals can execute intelligently. Rowlands also appreciates the capability of probabilistic reasoning in animal's cognition and deciphers: "Animals use cognitive maps to represent the possible locations of objects, assign subjective probabilities to the likelihood of an object being at a given location, and then update these probabilities as experience unfolds" (Rowlands 2019: 90).

> effect of other developments. But the history of the exercise of this capacity and its continual reapplication in criticism and revision of its products is not part of biology (Nagel 2013: 146).

A point that hails to my mind is to insist on ethics beyond the limit of biology; we may call it an unfair quest to ponder biology as the paradigmatic feature of morality. This unfair quest will make the goal for the animals to ensure the realm of morality without any access to self-knowledge, agency, and language-centric conceptual analysis of moral values. The behavioural and the theoretical patterns of ethics surge two different compatible models for humans and animals separately, and we have to accept this module. Nussbaum in her latest work *Justice for Animals: Our Collective Responsibility* (Nusbaum 2022) resonates with the capabilities approach of animals and writes:

> The Capabilities Approach can respond better than these other theories to the facts we now know about animal lives: about the amazing diversity of animal abilities and activities, about their capacities for valuing, for forming social networks, about their capacities for cultural learning, about friendship and love (Nussbaum 2022: 313).

I think a comparable assessment of behaviourism and cognitivism inculcates that the experiential behavioural progress of behaviourist theory is intended to overlook the cognitive ideology of cognitivism. For disdaining the museum myth, viz, minds, and mental processes, behaviourist stresses the exclusive account of the agent's external and observable behaviours. If we follow the behaviourist account, it looks appealing that the animals' behavioural patterns have certain intentional states with intelligence. We know that if we keep foods outside and in trap-boxes (case) for rats, in the first stage, the rats would favour taking the foods from outside since it looks safer for them. Next, they will observe the trap-boxes carefully, and if they find that one trap-box has only an entrance with no exit door, while the second one has entrance and exit doors together, the rat obviously chooses the second trap-box to get the food. So, doubtlessly they have the belief account, and their intentional intelligence guides their desires. The structural parts of beliefs have different components, but the components of the moral behaviour of animals' beliefs still remain inexplicable. The ethical pattern of the animal's behaviour looks puzzling since the module of this pattern is structured for humans and their linguistic-conceptual discourses. Here we need a meta-ethical juncture where the ethical and normative points have to be cognized through the parameter of moral motives and reasoning. Animals' sense of moral motives and their apparatus of getting involved in moral acts cannot be compared with human actions. Animals can act based on *some* moral reasons (moral reasons involving concern for others, taking responsibility for others, and rules followers, etc.). Frans de Waal writes: "Fatherhood changes the biology of men. New fathers experience a rise in oxytocin (the cuddle hormone) while their testosterone level drops. Men shift away from risk-seeking and mate-hunting toward a deeper commitment to their families. Their brains change as well. Neuroscience shows a more active and better-connected amygdala (emotional center) in the brains of men who are the primary caretakers in the family" (de Waal 2023). A somewhat similar approach to being able to live with, and care for families plus offspring is now known to be effective for animal

minds. Many stories of male apes, bonobos, and lions caring for their species or cubs can be found in various zoos and reserved forests.

Their position, I think is a threshold state for getting membership in the community of moral beings who have privileged based on basic capabilities like care, responsibility, individuality, etc. In the case of human moral engagement, we are stimulated by the typical moral theories and their diversified stances that could have conceptual and linguistic uses in our societal discourses. Still, animals' intentional apparatus may have the propensity to opt for an act following the moral consequences in their livelihoods. To find out its pragmatic consequences (utility or damage) tends towards a catastrophic understanding of animals' behaviours from an anthropocentric level.

References

Chakraborty, S. and Misra, S. 2022, 'Immanuel Kant', in Poff, D. and Michalos, A. (eds.), *Encyclopedia on Business and Professional Ethics*, Cham: Springer Nature.

Chomsky, N. 2006, *Language and Mind*, Cambridge: Cambridge University Press.

Crary, A. 2016, *Inside Ethics: On the Demands of Moral Thought*, Cambridge, MA: Harvard University Press.

Darwin, C. 1965, *The Expression of the Emotions in Man and Animals,* Chicago: Chicago University Press.

Dawkins, M.S. 1980, *Animal Suffering: The Science of Animal Welfare,* London: Chapman and Hall.

Dawkins, M.S. 1985, "The Scientific Basis for Asserting Suffering in Animals", in Singer, P. (ed.), *In Defence of Animals*, Oxford: Blackwell.

Descartes, R. 1970, *Discourse on Method*, Part iv, in *Descartes Philosophical Writings*, Anscombe, E. and Geach, P. (trans. & eds.), London: The Open University Press.

de Waal, F. 2023, "There Are Amazing Fathers in the Animal Kingdom", *Time*, https://time.com/6287477/fathers-in-the-animal-kingdom/ (last accessed October 2, 2023).

Flanagan, O.J. 1984, *The Science of the Mind*, Cambridge, MA: MIT Press.

Grueneisen, S., Duguid, S., Saur, H., and Tomasello, M. 2017, "Children, Chimpanzees, and Bonobos Adjust the Visibility of Their Actions for Cooperators and Competitors", *Scientific Reports*, 7, 8504.

Hauser, M. 2000, *Wild Mind: What Animals Really Think*, London: Allenlane Penguin Press.

Hunter, W.S. 1913, "The Delayed Reaction in Animals", *Behaviour Monographs*, 2, 6.

James, W. 1950, *The Principle of Psychology*, Vol. 1, New York: Dover.

Kant, I. 1988, *Kant Selections*, Back, L.W. (ed.), New York, London: Macmillan.

Korsgaard, C.M. 2006, "Morality and the Distinctiveness of Human Action", in Macedo, S. and Ober, J. (eds.), *Primates and Philosophers: How Morality Evolved*, Princeton: Princeton University Press.

Locke, J. 1993, An *Essay Concerning Human Understanding*, Yolton, J.W. (ed.), London: Everyman.

Lumsden, C. and Wilson, E.O. 1981, *Genes, Mind, and Culture: The Coevolutionary Process*, Cambridge, MA: Harvard University Press.

Malcolm, N. 1976, *Problems of Mind: Descartes to Wittgenstein,* London: George Allen and Unwin.

Nagel, T. 2013, *Mortal Questions*, New Delhi: Cambridge University Press.

Nussbaum, M.C. 2007, *Frontiers of Justice: Disability, Nationality, Species Membership*, Cambridge, MA: Harvard University Press.

Nussbaum, M.C. 2022, *Justice for Animals: Our Collective Responsibility*, New York: Simon and Schuster.

Putnam, H. 2022, "Thought and Language", in De Caro, M. and Macarthur, D. (eds.), *Philosophy as Dialogue*, Cambridge, MA: Harvard University Press.

Rowlands, M. 2009, *Animal Rights, Moral Theory and Practice*, New York, London: Palgrave Macmillan.

Rowlands, M. 2012, *Can Animals Be Moral?*, New York: Oxford University Press.

Rowlands, M. 2017, "Moral Subjects", in Andrews, K. and Beck, J. (eds.), *The Routledge Handbook of Philosophy of Animal Minds*, Oxon: Routledge.

Rowlands, M. 2019. *Can Animals Be Persons?*, New York: Oxford University Press.

Sartre, J.P. 1963, *Existentialism and Humanism*, Mairet, P. (trans.), London: Methuen and Co.

Scruton, R. 2000, *Animal Rights and Wrongs*, London: Metro Books.

Spinoza, B. 1982, *Ethics and Selected Letters*, Shirley, S. (trans.), Indianapolis: Hackett.

Strawson, P. 1977, *Individuals: An Essay in Descriptive Metaphysics,* London: Methuen.

Wilson, E.O. 1975, *Sociobiology: The New Synthesis,* Cambridge, MA: Harvard University Press.

Wittgenstein, L. 1953, *Philosophical Investigations*, Anscombe, G.E.M. (trans.), Oxford: Basil Blackwell.

Book Reviews

Christman, John, *Reconceiving Freedom from the Shadows of Slavery: Liberty in a Non-Ideal World*.
Cambridge: Cambridge University Press, 2025, pp. ix + 255.

John Christman's *Reconceiving Freedom* represents one of the most sophisticated attempts to articulate a positive conception of freedom in contemporary political philosophy. His work demonstrates remarkable philosophical precision and historical grounding. Christman develops his account with analytical rigor and clarity that makes complex philosophical concepts accessible, while his use of real-life examples from the history of slavery illuminates abstract discussions and demonstrates their practical significance.

Christman's sophisticated conception of freedom avoids three major problems that have historically undermined other positive accounts. First, he rejects the problematic notion of a "true self" existing independently from power dynamics, recognizing that our identities are formed through our relationships and social environments. Second, he distances his account from perfectionism, which ties freedom conceptually to particular forms of human flourishing. This allows freedom to function as an independent criterion for evaluating social arrangements. Third, he avoids moralizing freedom because he refuses to define freedom in terms of rights or actions sanctioned by a certain conception of justice, ensuring freedom can serve as an input into analyses of justice rather than being derived from prior moral commitments (142–143).

Central to his methodology is examining freedom as the antidote of slavery because slavery is "the most extreme version of any kind of tyrannical or oppressive circumstance" (34) and historically freedom emerges as a widely held social value directly in response to slavery.

Slavery operated first and foremost through the use of superior physical force and coercion. According to Christman, slavery was sustained and reinforced by three key mechanisms: *debilitation* (the systematic destruction of abilities needed to pursue valued projects), *social dislocation* (the forced disruption of stable social spaces through sale and movement), and *dehumanization* (the public racial degradation that denied full personhood). Since freedom is the antidote of slavery, freedom necessarily involves the necessary capacities to engage in meaningful social existence, the opportunity to develop self-governance, and public recognition as a person with full standing in the social order.

Christman argues that normative principles derived from overly idealized conditions distort the interests and perspectives of people living under systematic victimization. That is because people living under systematic victimization have interests centered on alleviating those conditions—concerns that he believes disappear when theorizing assumes idealized circumstances. He proposes a "fundamental constraint of political realism", whereby conceptions of freedom must guide action toward realizable goals (28). This constraint demands that if "freedom is impossible now and under any realistic medium-term future … we should reject (or put aside) that conception" (29).

But I see a tension between political realism and the conceptualization of freedom as slavery's antidote. Christman's claim that a conception of freedom should illuminate injustices while painting a graspable picture of an achievable future remains vague, as "realistic medium-term future" lacks a definition. Is this medium-term relative to a person's lifespan? North American slavery ex-

Argumenta 11, 1 (2025): 193–205
ISSN 2465-2334
DOI 10.14275/2465-2334/202521.boo
First published: 30 November 2025

tended from 1619 to 1885. Could someone in the XVII century conceptualize freedom in a way that recognized slaves as unfree persons without violating Christman's constraints of political realism? Consider the Quakers who petitioned against slavery in 1688. Neither they nor their grandchildren lived to witness the abolition of slavery. Should the Quakers have rejected or put aside their conception of freedom? If we follow Christman and reject conceptions of freedom simply because they are not currently realizable, we risk discarding valuable long-term ideals. Can we not distinguish between the conceptual clarity required to articulate long-term ideals and the more pragmatic frameworks needed for short- and medium-term goals?

Christman presents his positive conception of freedom as follows:

> A person is free to act or engage in an activity [or practice] or way of life to the extent that she has the opportunities, capabilities, and social support to engage in the self-governing pursuit of those activities that she can endorse from the perspective of her practical identity, and she is socially recognized as eligible for these pursuits (81).

This is the definition of what Christman calls *scalar freedom*. But Christman also introduces the notion of *basic freedom*, defined as a sufficient degree of scalar freedom. A person possesses basic freedom iff she has an adequate range of opportunities, capabilities, and social support to pursue socially constituted practices "that she can endorse from the perspective of her practical identity; and she is socially recognized as being a self-governing agent who can engage in such pursuits" (82). One possible interpretation of this formulation is that Christman may be blurring the line between two distinct questions: what positive freedom *is*, and what social conditions are necessary to *guarantee* or *support* it. In other words, he may be conflating the concept of positive freedom itself with its enabling conditions—such as recognition, capabilities, and institutional support—which raises the question of whether these should be considered definitional features of freedom or simply instrumental to it.

Christman maintains that "a free society", that is, a society that values freedom, should ensure its members enjoy basic freedom (106). However intuitive this distinction appears, the concept of basic freedom remains vague because it is unclear what constitutes an "adequate range" or what degree of scalar freedom suffices for a society to qualify as free. One possible way to interpret this is to treat "a free society" as a kind of range property: that is, a binary designation supervening on a scalar property (i.e., basic freedom of individuals). From this perspective, basic freedom is scalar, varying in degree, but a society counts as *free* only if each of its members possesses at least a minimum threshold of that freedom. This helps clarify how Christman might be using the term "free" at the societal level while maintaining a scalar framework for basic freedom of individuals. For analytical clarity, I now turn to examine the four components of scalar freedom in greater detail.

First, what does freedom pertain to? Christman argues that freedom pertains to *actions*—both basic (performed directly) and non-basic (performed by performing another action)—*activities* (sets of actions organized around purposes within socially understood scripts), *practices* (activities structured by normative standards), and *ways of life* (sets of practices connected to identities). These categories are inherently temporal, interpersonal, and socially constituted. For the

sake of simplicity, I refer to all these collectively as *socially constituted practices*. They derive meaning from shared understandings that cannot be reduced to mere aggregations of actions. By contrast, according to pure negative and republican conceptions, freedom only pertains to actions: one is free or unfree solely with respect to actions.

To support his claim that freedom does not pertain only to actions, Christman offers three arguments. First of all, identical sets of actions can have different meanings depending on social relations and institutional contexts. When *A* enters a sanctified church and kneels before an ordained priest, he "goes to mass". When *B* performs identical physical movements in a deconsecrated church with a faux priest, she does not "go to mass". Though both persons have the negative freedom to perform the same actions, only *A* engages in the activity of attending mass (94–96). Since humans aim to engage in socially meaningful pursuits rather than isolated actions devoid of context, a conception of freedom should enable us to evaluate whether individuals are free to engage in socially constituted practices.

Moreover, by drawing on Bratman's work, he argues that humans are "planning creatures" whose actions exist within temporally extended plans and self-governing policies (99). Thus, human actions are components of hierarchically organized plans directed toward goals. These plans are "partial". They allow various paths to achieve objectives. Crucially, actions are almost always socially embedded. They gain meaning from institutional structures and shared understandings. Thus, actions can only be understood through narratives that situate them within socially constituted practices. A plan to write a philosophy paper, for instance, only makes sense within a socially constituted academic practice with specific norms (100). If freedom pertains to actions, and actions are embedded in plans and socially constituted practices, then freedom must also pertain to these practices.

Christman also argues that the concept of a *basic action* is "unstable", as any action can be infinitely subdivided (97). Pure negative freedom theorist Carter has suggested resolving this *problem of infinite subdivision* by dividing space, time, and matter (the human body and its related objects) into equally sized, finite units.[1] Other prominent pure negative theorists have endorsed this approach.[2] However, Christman does not address this debate.

Furthermore, Christman argues that according to pure negative theorists, freedom is valuable because it enables agency[3] and preserving freedom shows respect for deliberative capacity.[4] He believes that these claims presuppose a richer conception of agency than mere choice-making, which, in Christman's view, pure negative freedom theorists implicitly rely on. Agency, on this richer conception, belongs to temporally extended persons whose actions reflect their practical identities, that is a person's "ongoing self-concept that structures her evaluative perspective, a description of herself under which things have value"

[1] Carter, I. 1999, *A Measure of Freedom*, Oxford: Oxford University Press, 184.
[2] Carter, I., Steiner, H. 2022, "Freedom Without Trimmings: The Perils of Trivalence", in Kurki, V.A.J., McBride, M. (eds.), *Without Trimmings: The Legal, Moral and Political Philosophy of Matthew Kramer*, Oxford: Oxford University Press, 239–67, 254; Kramer, M.H. 2003, *The Quality of Freedom*, Oxford: Oxford University Press, 380.
[3] Carter 1999: 58.
[4] Kramer 2003: 241.

(78). Since these identities manifest through social roles (e.g., parent, academic, spouse), valuable freedom must encompass socially structured activities. If freedom were limited to actions divorced from social meaning, its connection to identity would dissolve, undermining the very basis for its value that pure negative theorists themselves recognize.

A significant weakness emerges in Christman's discussion of freedom's non-specific value. Carter argues that freedom is valuable as such, that is, it has a *non-specific* value because its value cannot be reduced to the value of specific actions a person is free to perform.[5] For example, even freedoms we cannot currently imagine valuing remain valuable due to our ignorance about future desires. Christman claims that his account captures this non-specific value. He asks us to imagine a religious person who values *the freedom to sin* because it provides the opportunity to resist temptation and increase piety (125). According to Christman, this shows that the freedom to sin has non-specific value as its value is not derived from the act of sinning itself. However, this example in fact demonstrates a kind of *specific* value, *viz*, a specific constitutive value of the freedom to sin. The person values the freedom to sin because it is conceptually necessary for being pious. Being a pious person necessarily involves freely choosing not to sin rather than lacking the opportunity altogether. Thus, the value of this freedom is specific because it is explained by its constitutive role in piety.[6]

Second, what are the conditions of freedom and unfreedom? Some theorists define freedom as the absence of prevention by others. Christman considers such a conception "empty" because it ignores whether individuals can carry out the unprevented actions (114). For him, freedom has four individually necessary and jointly sufficient conditions: (i) the absence of preventive obstacles imposed by others; (ii) possession of relevant capabilities (abilities and resources); (iii) social recognition as eligible to engage in the pursuits; and (iv) the presence of socially constituted practices that the agent can endorse from the perspective of her practical identity. What are the relevant capabilities mentioned in (ii)? According to Christman, not all capabilities are necessary for freedom. Any inability produces unfreedom when it prevents socially constituted practices that align with one's practical identity. Thus, only those capabilities enabling socially constituted practices one values given one's practical identity are necessary for freedom. For example, as an atheist analytical philosopher, my lacking the capabilities to engage in practices requiring unquestioning religious obedience does not constitute unfreedom. Given this practical identity, freedom does not pertain to such practices. If my practical identity were to change, the relevant capabilities for freedom would change accordingly.

In addition, Christman holds that an inability constitutes unfreedom only if either: (a) it is caused by others' actions, or (b) it can *feasibly* be removed by others. This means that it is possible for a person to be unable and nevertheless not unfree to do something. In other words, Christman draws a distinction between unfreedoms and mere inabilities (119). However, point (b) introduces significant vagueness in determining which obstacles are removable because determining which obstacles are *feasibly* removable requires *moral judgments* that different ethical frameworks might answer differently. This is because feasibility does not

[5] Carter 1999: 34.

[6] For a relevant discussion, see Huseynli, I. Forthcoming, "The Ability Independent Value of Freedom", *European Journal of Political Theory.*

simply refer to what is physically or technically possible. For example, many inabilities of disabled individuals, such as the inability to walk due to limb loss, could be addressed if sufficient resources were directed toward producing prosthetics, subsidizing their distribution, or advancing prosthetic technology through research. Similarly, many chronic or rare diseases that limit a person's abilities could potentially be cured if adequate funding were allocated to medical research. Many inabilities experienced by disadvantaged groups, such as women, LGBTs or lower-caste individuals could be removed if communities were willing to reform their cultural or religious doctrines. Even something as extreme as my inability to walk on the Moon or Mars could be overcome: with enough investment, I could be trained and sent to the Moon using existing technology, and with continued funding and research, travel to Mars may become possible in my lifetime. Whether such efforts are deemed feasible depends on prevailing moral (and/or political) priorities. Christman's criterion of feasibility introduces a normative dimension into what might otherwise seem like a straightforward descriptive distinction. Christman explicitly acknowledges but does not eliminate this vagueness (120).

Third, what are the conditions of self-government? Self-government stands at the core of Christman's conception, as it enables individuals to pursue activities that align with their evaluative perspectives. Self-government consists of two essential conditions. First, *the capability condition* requires that individuals possess the personal capabilities necessary to form reflective intentions. These include basic subsistence capabilities such as adequate nutrition, shelter, clothing, and health resources as well as education and skills relevant to one's social context. The specific capabilities required will vary across different societies, but the overarching criterion is that they enable a person "to formulate, deliberate upon, and act intentionally on the basis of values that can guide her choices of a way of life" (136). Second, *the reflective affirmation condition* requires that individuals can see their socially constituted practices as expressions of their practical identity. This does not demand that one has created or controls all aspects of one's life. Instead, freedom requires that one can affirm one's condition by seeing one's choices as intelligible extensions of one's evaluative orientation. When engaging in socially constituted practices that meet this condition, one's actions recursively affirm one's identity, creating "a kind of ongoing self-affirmation" (151).

Fourth, why is social recognition necessary for freedom? Social recognition is essential because it provides the public validation required for meaningful participation in socially constituted activities. Without such recognition, individuals cannot be self-governing regardless of their internal capabilities, as their activities lack social legibility and support. Christman identifies two distinct vectors of recognition. *Second-personal recognition* operates at the interpersonal level when others acknowledge an individual's needs and identity in direct interactions. This occurs for example when people use someone's preferred pronouns or address their ethnicity respectfully. The absence of this recognition manifests as discrimination or social marginalization in face-to-face encounters. *Structural recognition* functions at a broader societal level through the generally shared standards of achievement and social standing expressed in formal and informal institutions. This includes the shared meanings of socially constituted practices and what counts as accomplishment within institutional arrangements. These two vectors often interact but remain conceptually distinct. One might enjoy structural recognition while occasionally facing interpersonal disrespect.

I think Christman's claim that freedom pertains only to socially recognized practices renders his account conservative. It cannot accommodate freedom infringement claims regarding practices unrecognized by society. Consider marriage. Traditionally defined as a union between a man and a woman, Western societies later redefined it as a union between two adults. Christman uses this example to show that practices can change when advocates convince society that new changes align with fundamental principles. However, his account implies that before same-sex marriage was legally recognized, homosexuals could not claim they were *unfree* to get married because freedom did not pertain to same-sex marriage at all. At best, they could argue that recognizing same-sex marriage would increase their overall freedom by adding to the practices they could engage in, or that they lacked basic freedom because they could not live according to their practical identities. This weakens the normative force of freedom-based arguments for social change. Appeals to *freedom infringement* are powerful because they imply disrespect and suggest moral responsibility for this infringement. Yet Christman's account disarms those advocating for changes in practices by preventing them from claiming unfreedom in contexts where the practice lacks social recognition.

Overall, *Reconceiving Freedom* is a powerful and original defence of a positive conception of freedom. Yet its reliance on social recognition in the definition of freedom, and its insistence that conceptions of freedom must be realizable in the medium term, present significant theoretical challenges. These tensions do not undermine the value of the book but rather underscore the enduring difficulty of articulating a positive conception of freedom that is both normatively ambitious and politically grounded.

University of Pavia ILKIN HUSEYNLI

Iaquinto, Samuele & Torrengo, Giuliano, *Fragmenting Reality: An Essay on Passage, Causality, and Time Travel.*
London: Bloomsbury Academic, 2022, pp. x + 192.

Realism about tense is roughly the view that some of the facts that fundamentally constitute reality are irreducibly tensed. It is something like this view that John McTaggart (1908) thought to be contradictory, but most philosophers remain convinced that his original argument for this conclusion unequivocally fails. But a recent reconstruction of the argument due to Kit Fine[1] demonstrates that realism about tense does indeed lead to a contradiction if combined with the following three principles about reality:

NEUTRALITY
No time is privileged, the tensed facts that constitute reality are not oriented towards one time as opposed to another.

ABSOLUTISM
The constitution of reality is an absolute matter, i.e., not relative to a time or other form of temporal standpoint.

COHERENCE
Reality is not contradictory, it is not constituted by facts with incompatible content.

[1] Fine, K. 2005, "Tense and Reality", in *Modality and Tense: Philosophical Papers*. Oxford: Oxford University Press, 261–320, §§3–5.

Consider some tensed fact—say, your being asleep (f_1, for short). Given NEUTRALITY, reality is not exclusively constituted by the facts obtaining at this particular moment. Thus, reality is also constituted by tensed facts obtaining at some other time, such as the fact that you are awake (f_2). Given ABSOLUTISM, the conflict between these facts cannot be resolved by saying that, relative to the one time, reality is constituted by f_1 yet not by f_2, and relative to another time, it is constituted by f_2 yet not by f_1. Reality is supposed to be constituted, absolutely speaking, by both of these facts, and thus COHERENCE will be violated.

The by far most popular realist reaction to this conundrum has been to reject NEUTRALITY and to endorse what we may term *elitism* about the constitution of temporal reality. According to this view, it is only the tense facts obtaining at an absolutely privileged moment of time, the one that is absolutely present, that get to constitute reality. At first glance, this may seem less revisionary than the other two options available to the realist, but Fine[2] powerfully argues that this standard form of realism fares far worse than the non-standard varieties abandoning either ABSOLUTISM or COHERENCE when it comes to accommodating the intuitive sense in which time passes, delivering a plausible semantics for utterances of tensed sentences, and squaring realism about tense with special relativity. There is, thus, ample reason to prefer *egalitarianism* over elitism about constitution, and to view the various tensed facts obtaining at different times as making an equal contribution to the constitution of reality.

Egalitarian, non-standard realists still face a choice between ABSOLUTISM or COHERENCE. Fine has dubbed the view that rejects the former *external relativism*, and the view that rejects the latter *fragmentalism*. Roughly, external relativists hold that the question of what facts constitute reality makes no sense unless one understands it as being posed from within a temporal standpoint—there is no such thing as reality's just *being* a certain way, absolutely speaking. And the correct answer to the temporally relativised questions may vary from one standpoint to another. Fragmentalists, by contrast, do not dispute the legitimacy of posing the question from an *atemporal* standpoint. It is just that there is no *coherent* answer to it—it is not so much the question as the answer that does not make sense. There is a certain way that reality is, absolutely speaking; yet that way is incoherent, as it involves facts with incompatible content. I must confess that I find fragmentalism significantly more radical and, at least initially, far less plausible than external relativism—in fact, if giving up COHERENCE were the only way of upholding realism about tense, I would have thought that the price is not quite right. But many (or perhaps even most) seem to disagree. Fine (2005: §11) himself has argued that fragmentalism has several advantages over relativism; accordingly, while a number of philosophers have taken an interest in fragmentalism, the *other* non-standard view has been largely ignored.[3]

In any case, we now have the first book-length treatment of fragmentalism: Samuele Iaquinto and Giuliano Torrengo's *Fragmenting Reality*. This is a really impressive book—a most welcome contribution to the growing literature on fragmentalism and definitely a must-read for anyone with an interest in the topic. After some initial stage-setting in an introductory chapter, the authors put

[2] Fine 2005: §§7–10.

[3] Largely, but not completely: see Eker, B. 2023, "Perspectivalism about Temporal Reality", *Synthese*, 202 (2), 1–29; Eker, B. Forthcoming, *Time and Temporality in Perspective*. London: Palgrave Macmillan.

forth in the first chapter what is the most in-depth investigation of the fragmentalist project to date, one that includes very helpful and detailed comparisons between their own particular brand of fragmentalism, Flow Fragmentalism (*F-Fragmentalism*, for short), and various other fragmentalist approaches previously presented. The subsequent chapters are dedicated to showcasing the explanatory potential of the view, by examining such topics as the open future, the metaphysics of causation, physical relativity, and time travel through the lens of F-Fragmentalism. Throughout, the book is clearly written, rich in argument, and packed with nuanced discussions of various important details of the relevant subjects. This last point is among the reasons why this is such a rewarding book; but it also means that, given the limitations of space, it is impossible to do proper justice to the authors' wide-ranging exploration of the fragmentalist landscape. Hence, I will have to confine myself to a few basic aspects of F-Fragmentalism in my critical remarks below.

Any form of fragmentalism worth its name comes equipped with a temporally absolute and tenseless notion of constitution (in accordance with ABSOLUTISM); after all, this is what is supposed to distinguish this variant of non-standard realism from the other, relativist one. In this respect, the F-Fragmentalist follows suit. Yet one distinguishing feature of F-Fragmentalism is that Iaquinto and Torrengo also introduce a further, "irreducibly relative" notion of obtainment (21). They emphasise quite a few times that this is no idle terminological preference, writing that "the distinction between constitution and obtainment is particularly important for the theory" (2). According to this picture, while reality is temporally absolutely constituted by both your being asleep and your being awake, neither of these facts obtain *simpliciter* but rather only relative to distinct fragments of reality.

Far be it from me to criticise the F-Fragmentalist purely on the grounds that she embraces an irreducibly relative notion of obtainment—indeed, I am in general convinced that this is the way to go for the non-standard realist. However, the way Iaquinto and Torrengo appeal to the notion raises several difficulties. Consider this passage, for instance:

> [T]he distinction between (absolute) *constitution* and (irreducibly relative) *obtainment* is between two *ways* facts relate to reality in the fragmentalist's picture, none of which can be explained in terms of the other. F-Fragmentalist is thus a form of pluralism […] in the sense that it comes with two distinct metaphysical commitments with respect to how facts relate to reality. Neither of them is explanatorily more fundamental, and both are equally metaphysically perspicuous, in their own ways (22).

But if there are indeed these two fundamentally different ways in which facts relate to reality, then the Finean principles must apply *separately* to *each*; and given this, F-Fragmentalism seems to involve a rejection of *both* ABSOLUTISM *and* COHERENCE: the former is abandoned with respect to obtainment, and the latter with respect to constitution.[4] The basic problem with this should be obvious: any view that rejects *two* of the Finean principles seems to be ideologically less parsimonious than a view that only rejects one of them.

[4] Loss considers something like this view, dubbing it "Fragmentalist Relativism", in Loss, R. (2017) "Fine's McTaggart: Reloaded", *Manuscrito*, 40, 209–239, 219.

On closer inspection, however, this kind of "egalitarian overkill" does not seem to be what Iaquinto and Torrengo are after. One factor that complicates the issue is that it remains rather obscure what distinct ways constitution and obtainment are supposed to be for facts to relate to reality. It seems clear that what it is for a fact *f* to constitute reality is for reality to be the way that *f* specifies: given that your being asleep constitutes reality, your being asleep is a way that reality is. The question is what distinctive role "irreducibly relative obtainment" is supposed to play in this picture. Iaquinto and Torrengo hold that "obtainment relative to a fragment tells us how what is the case organizes into fragments" (25). But obtainment relative to a fragment may "tell" us how the facts divide into fragments without there being any irreducible fragment-relativity in how facts relate to reality. For Iaquinto and Torrengo characterise the fragments relative to which facts are supposed to obtain as "maximally coherent collections of tensed facts" (47). Indeed, they claim that "[w]hat a sentence such as ['Relative to fragment F, φ'] states is simply that the fact that φ is part of a certain collection of facts (which cohere with one another)" (46). But this suggests that obtainment relative to a fragment is to be analysed as follows: what it is for a fact *f* to obtain relative to some fragment F just is for *f* to be, absolutely, a member (or part) of F. In other words, obtainment relative to fragments is analysed in terms of absolute membership (or parthood), which, in turn, will presumably be analysed in terms of absolute coherence relations holding between facts. Given this, there seems to be no fundamental fragment-relativity of any kind in the way in which your being asleep or your being awake relate to reality: both of these facts constitute reality absolutely, and hence things absolutely (yet incoherently) are both these ways. But given that these facts do not cohere (absolutely) with each other, they belong (again, absolutely) to distinct maximally coherent collections of facts.

In brief, *pace* Iaquinto and Torrengo, F-Fragmentalism does not seem to feature any genuinely irreducible fragment-relativity. This confirms that the view is a more orthodox version of fragmentalism, after all. On the other hand, the fundamentally relative notion of obtainment appears to be vital for the F-Fragmentalist accounts of the passage of time and the open future, and my worry is that much of the plausibility of those stories will evaporate once the F-Fragmentalist notion of relative obtainment is revealed as bogus.

There is certainly much more to be said about the above issues, but let me close by briefly voicing a more fundamental difficulty with F-Fragmentalism, one that arises for any view that attempts to combine realism about tensed facts with the idea that the constitution of reality by facts is a temporally absolute, tenseless matter.[5] The difficulty is, in a nutshell, that this combination of views seems inconsistent. For if the constitution of reality by facts is a temporally absolute, tenseless matter, then there must be a certain way that reality temporally absolutely, tenselessly is—precisely the way specified by the facts that do constitute it in this temporally absolute, tenseless manner. But given this, how could any of those facts that are involved in the tenseless constitution of reality be themselves *tensed*? If whether or not your being asleep constitutes reality, whether or not reality is that way, is a tenseless matter and does not depend on anything temporal at all, how could it still be a tensed matter whether or not you are

[5] For an extensive discussion of the argument, see Eker, B. 2023, "Realism about Tense and Atemporality", *Synthese*, 202 (5), 1–25, and Eker, B. Forthcoming, *cit.*, Ch. 4.

asleep? How could it be that, while constitution is a fundamentally tenseless, temporally absolute matter, the reality that thereby gets constituted is fundamentally tensed? Given that whether or not the facts constitute reality is a tenseless matter, what about those facts could it be that makes them tensed rather than tenseless?

The picture that emerges from a fragmentalist metaphysics depicts reality to be all over the place—there is a way of specifying how things are, *once and for all*, yet it does not quite make coherent sense. I think that this constitutes, on its own, a fairly steep cost. But it is considerations along the above lines that I take to be really decisive against the project of developing non-standard realism about tense through the fragmentalist route—fragmentalists forms of *realism* are misguided because their commitment to ABSOLUTISM seems to contradict the tensed character that the facts are supposed to have. As far as I can see, F-Fragmentalism constitutes no exception to this. Still, none of these doubts about the ultimate viability of a fragmentalist approach to tense realism detracts from the philosophical merits of this book, which has much to offer to anyone thinking about the fundamental nature of time and temporal reality.

University of Hamburg BAHADIR EKER

Salami, Minna, *Sensuous Knowledge: A Black Feminist Approach for Everyone*.
London: Zed Books, 2020, pp. v + 190.

It was through a writing style that combines personal reflection, cultural criticism, and political analysis—anchored in a Black and transnational feminist perspective—that Minna Salami's voice became distinguished in contemporary feminist thought. Her essays and lectures move fluidly between reflections on gender, race, philosophy, and art, marked both by a lyrical tone and by a commitment to epistemological transformation. *Sensuous Knowledge* continues and consolidates the path first outlined with the MsAfropolitan.com platform. By bringing together the intimacy of an almost diaristic writing style with theoretical insight, Salami crafts a work that places Black feminism at the center of understanding the world, employing a language that is accessible and sensitive, yet intellectually rigorous.

The introduction sets the tone for the book. Salami recounts the story of two explorers in search of a legendary mountain. The first returns frustrated, claiming to have found a barren and lifeless place, while the other, ecstatic, describes encountering a lush, verdant landscape. Through this metaphor, Salami aligns herself with the second explorer, offering us a renewed and enriching perspective on knowledge. Her analytical lens is rooted in an "Africa-centered black feminist angle" (2), countering the dominant Eurocentric and patriarchal—what she calls "Europatriarchal"—vision she critiques as one-dimensional. Her aim, however, is not merely to criticize the Europatriarchal view but to highlight and value the narratives it has obscured. Inspired by the integration of spirit, joy, and reciprocity with nature, the chapters of this book can be seen as portraits of the other side of the mountain, inviting readers to reflect, to imagine, and to formulate new ways of knowing.

One of the book's strengths is how Salami defines knowledge. In *Of Knowledge*, she notes that she does not intend to engage in academic epistemology debates but prefers to draw from the Yoruba creation myth to illustrate her understanding. According to this mythological narrative, knowledge presents a duality: *ogbon-ori* and *ogbon-inu*, "knowledge of the head" and "knowledge of the gut", respectively, with true knowledge integrating both forms. However, as Salami diagnoses, Western modernity has privileged only one form, *ogbon-ori*, marginalizing embodied and emotional experiences. Thus, she advocates for the integration of *ogbon-ori* and *ogbon-inu*, proposing a "sensuous knowledge" as an expansion of modern knowledge. Inspired by Black feminists, she calls for the incorporation of art, poetry, and love to enhance our emotional intelligence; critical of a neutral Eurocentric science, she proposes the need for a more ethical and inclusive science; aware of colonial illusions, she highlights African knowledge that promotes interconnection between humans and nature. Her goal is to challenge current power structures in search of a new approach to knowledge that values both the measurable and the immeasurable.

From the third chapter onward, Salami delves into areas of knowledge through narratives involving her private life stories. In *Of Liberation*, she intertwines her personal experiences with the politically engaged music of Ms. Lauryn Hill and other artists challenging the cultural establishment. Inspired by the transformative power of music, Salami shares her own journey into writing as an act of liberation from an education marked by violence and prejudice. Writing is the expression Salami found in her quest for a more authentic life. For her, societal transformation begins with personal transformation, seeking self-knowledge and the strengthening of our sense of self. In a world filled with injustices and inequalities, music and art provide a vital space for critical thinking and emotional expression. Quoting Ms. Lauryn Hill, Salami reminds us that seeking freedom is not easy, but a journey that requires persistence: "But no! Not forever like this will I be, here are my hands that can really make me free" (59).

In *Of Decolonization*, Salami describes the complexity of dealing with colonial and traditional names by sharing her relationship with Anthony, a Ghanaian with a British name, and how this experience leads us to reflect on the impacts of colonial heritage on our minds, bodies, and souls. Referencing the thoughts of Kenyan novelist and intellectual Ngũgĩ wa Thiong'o, she discusses the pitfalls of mental colonization. To decolonize, in her view, means recovering territories and reconstructing social structures that free us from mental prisons. With this, Salami advocates for a comprehensive transformation of consciousness, one that clearly breaks away from patriarchy and racism and embraces diverse cultural expressions, even those remnants of colonialism, like the language in which she writes this book: English.

In the next chapter, Salami invites us to rethink identity as a source of joy and liberation. She begins this conversation by describing Peter Paul Rubens' painting "The Four Philosophers", which portrays the stereotype of the ideal philosopher: four white, middle-aged European men. This painting is no coincidence; it symbolizes how the ideas, values, and concerns of this group dominate our institutions, culture, economy, and politics. Salami imagines an alternative version of this scene, where four Black women occupy the men's place. How would these women's philosophical analyses influence our collective imagination? For her, the true weapon against this oppressive system is pleasure and joy. "Joy", in her words, means, "freeing yourself from predefined notions of

identity" (78). In this sense, the pursuit of a happy life is a weapon that challenges, for example, notions of femininity that predestine women to eternal suffering, to an incomplete and dependent identity. Therefore, Salami imagines identity as a compass guiding us towards joy, helping us find new perspectives to address injustices.

"One is not born, but rather one becomes, black" (89). With this adaptation of Simone de Beauvoir's famous phrase, Salami begins the chapter *Of Blackness*, showing how blackness is also a socio-historical construction connecting all "people of Black African heritage" (88). However, the concept of blackness is not perceived the same way by Black people in all nations. In Nigeria, where Salami spent her childhood, identifying as Black means aligning with what is considered inessential. It was only upon encountering whiteness and its racism when she moved to Sweden that Salami realized her Black identity. Thus, Salami critiques the reluctance of many Africans to adopt the Black identity. For her, ignoring blackness means losing a crucial connection with the African diaspora and the struggle against imperialist exploitation. Therefore, she advocates that people on the African continent value blackness as a way to preserve the connection of their historical legacy with their present, uniting the fight against white supremacy with the fight against capitalist exploitation of the African continent. Blackness is "the harmony that gels the past, present, and the future of African heritage people" (100).

Of Womanhood is a chapter divided into phases of the moon, each representing a different stage of the female journey, from struggle and survival to trauma and healing, culminating in the assertion of a language that reflects autonomy. It is moving how Salami narrates her own experience of socialization as a woman, revealing a process marked by fear and subordination. Her critical view highlights the persistence of anti-woman institutions and knowledge, even in the face of women's achievements worldwide. Within the Europatriarchal logic, being a woman means occupying a place of lack—lack of masculinity—and consequently of submission and self-loathing. Thus, resonating with ideas from intellectuals like Simone de Beauvoir, Shulamith Firestone, Nina Power, Iris Young, African voices like Franz Fanon and Oyeronke Oyewumi, and radical lesbians, Salami advocates for the construction of a new language to think about womanhood, one informed by sensuous knowledge, both through the mind and intellect and through lived passions and emotions. Instead of pink, Salami claims blue as the color of womanhood, "the color of wholeness and unity, it represents the confluence of the sky and the seas where the earth becomes one" (118).

In *Of Sisterhood*, Salami inspires us to seek a feminism that recognizes and celebrates differences. Addressing the polarization within the movement, Salami highlights how differences in class, race, nationality, religion, and sexual orientation have divided women. However, she believes feminism must be both a struggle against patriarchy and a fight for female solidarity. With this in mind, she reconstructs the history of the feminist movement from the margins, bringing the narratives of Black and lesbian women to the center and their struggles against the hegemony of white feminism. She argues that, unlike many ideological movements, feminism welcomes criticism and improves itself. Since the publication of the famous anthology "Sisterhood is Powerful",[1] more inclusive

[1] Morgan, R. 1970, *Sisterhood Is Powerful: An Anthology of Writings from the Women's Liberation Movement*. New York: Vintage Books / Random House.

and comprehensive works have been produced, reflecting an evolving movement seeking solidarity among women worldwide. Salami emphasizes that collaboration between white and Black feminists is essential. Echoing Toni Morrison's view, she stresses the need to rid ourselves of both hate and guilt. For her, one of feminism's greatest challenges today, as it always has been, is understanding that we are "one" body.

In the next chapter, the flow of the river guides her writing. In *Of Power*, Salami warns of the importance of rethinking patriarchal definitions of power, which have associated power with the state and institutions; as synonymous with domination, violence, and coercion; and as something quantifiable. Instead, inspired by Hannah Arendt's visions, who defines power as an act in concert, and Mary Parker Follett's proposal of "power with" rather than "power over" others, Salami explores power as an action. For her, power is a vital force, reminiscent of the Yoruba's ashe, the Hindus' prana, and the kami of Shinto worshipers in Japan. Through this movement, she compares power to gravity for rivers, a force that helps us navigate the twists and turns of life towards self-actualization. Inspired by this metaphor, Salami believes rivers teach us to revolt, but not violently, controllingly, or quantifiably. The river teaches us the strength of the collective and the individual, a force that sustains all living species without needing control mechanisms. Power simply "is".

It is fitting to conclude the book with the chapter *Of Beauty*. In it, Salami describes three types of beauty. The first facet of beauty is political, originating from the Bible, especially in the representation of Eve in Genesis. Salami reveals that behind the imposition of these aesthetic standards lies a political agenda that exalts male desire for power through the submission of the female body. The second facet of beauty is "artificial". Here, she explores how aesthetic procedures, such as skin bleaching, relate to the notion of "worldliness", i.e., the pursuit of conformity to the standards of a global and cosmopolitan community, often associated with whiteness and aligned with the interests of a capitalist society. In this context, what could be genuine beauty? Salami offers us some insights: beauty is anarchic, a rebellion against conformity; it consistently reflects a unique cultural experience; it is our personal secret, something we nurture within our internal world of fantasy and imagination; it is eternal and glorious from the dawn of time; and ultimately, it is action, acknowledging that forms of oppression cannot embody true beauty. For Salami, true beauty is the daily practice of being always in love with oneself and life, something that goes hand in hand with the political agendas she has explored throughout the book.

Sensuous Knowledge is a work that both challenges and enriches our understanding of knowledge. One should not expect a conventional academic book, for its purpose is precisely to rethink traditional ways of knowing, integrating theory with intuition, the mind with the body.

Federal University of ABC - UFABC MICHELE BONOTE

Advisory Board