EUROPEAN JOURNAL OF ANALYTIC PHILOSOPHY

UDC 101 | ISSN 1845-8475

Vol. 21, No. 2, 2025

EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDC 101
ISSN (Print) 1845-8475
ISSN (Online) 1849-0514
https://doi.org/10.31820/ejap

Open access

Publisher, Editorial office

University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy
Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia (EU)
Phone: +385 51 669 794
E-mail: eujap@ffri.uniri.hr
Web address: https://www.eujap.uniri.hr

The publication of the journal is financially supported by the Ministry of Science, Education and Youth of the Republic of Croatia under a programme agreement with the University of Rijeka.

TABLE OF CONTENTS

BOOK SYMPOSIUM

RESEARCH ARTICLES

PRÉCIS OF MADNESS: A PHILOSOPHICAL EXPLORATION

Justin Garson[1]

[1] Hunter College and The Graduate Center, City University of New York, USA

Received: 29/04/2025 Accepted: 25/05/2025

ABSTRACT

The following is a short synopsis of the book *Madness: A Philosophical Exploration*. It provides an overview of the book's core distinction between madness-as-dysfunction and madness-as-strategy, and enumerates four benefits of relying on this conceptual framework: for history, philosophy, Mad Pride, and treatment.

Keywords: psychiatry; mental disorder; madness-as-dysfunction; madness-as-strategy.

Correspondence: jgarson@hunter.cuny.edu

I see *Madness* less as a monograph, complete with a proper thesis and argument structure, and more as a series of conceptual exercises. Its goal is neither to instruct the reader about the intellectual history of psychiatry nor to demonstrate a philosophical thesis about the nature of mental illness. Rather, it aims to induce a certain perspective shift in the reader. Hence the book's warning: "There is no synopsis or abbreviation or précis that can possibly serve as a substitute for reading the book". (Garson 2022, 6) It's more like a machine that either does its job or fails to work the way it should.

The somewhat peculiar aim of the book also problematizes the very idea of a book symposium. As traditionally understood, a book symposium consists of a series of critiques alongside an extended defense. (It's notable that book symposia are sometimes called "author-meets-critics" sessions.) Such symposia evoke the idea of an author confronted, face-to-face as it were, by a series of open-minded but skeptical readers, readers who are prepared to indicate the ways in which the book falls short of its promise, by, say, exposing fallacious arguments or highlighting unacknowledged alternatives. As the book is less of a monograph than a machine, it imposes a different set of questions on the symposiast—not *Did Garson convince me of his thesis*? but *Is the book a good tool*? *Are there better tools for the job*? *Is it a worthwhile job in the first place*?

With those qualifications in mind, however—and given the book's explicit rejection of the idea that it could be encapsulated by a précis—I will describe as briefly as possible the purpose of the tool, the manner of its operation, and the benefits of its deployment.

Here are some background beliefs that I hold. The history of psychiatry, or better, the history of madness, can be seen as a clash or confrontation between two major paradigms or worldviews. Most major theorists of madness can be comfortably seen as accepting one or the other of these systems, and most of the major disputes and paradigm shifts in the study of madness can be usefully viewed through this lens. The clash that I envision, however, is not equivalent to any of the clashes that we are used to hearing about, such as that between biological and psychological worldviews or that between reductionistic and holistic worldviews. Rather, it is between *dysfunction-centered* and *function-centered* worldviews, or what I call *madness-as-dysfunction* and *madness-as-strategy.*

Madness-as-dysfunction and *madness-as-strategy* are not meant to describe theories about madness; they are meant to describe psychological tendencies on the part of researchers, tendencies to "approach" madness in a certain way. The core idea behind *madness-as-dysfunction* is that when somebody is mad (construed broadly to include things like extreme low mood, panic attacks, hearing malicious voices, having strange beliefs that other people don't accept, being such a "difficult person" that it's hard to hold down a job, and so on), it's because something inside of them, in their mind or brain, is not working the way it is supposed to. Seen this way, madness (or its various expressions: malicious voices, low mood, etc.) is seen as a "symptom" of a "disorder". It is the kind of thing you might see a doctor to diagnose and treat, perhaps with medication or therapy.

(*But wait a minute, you've just described psychiatry itself!* Exactly—my point is that *madness-as-dysfunction* has become so entrenched in psychiatric thought that it's difficult to recognize as a distinct paradigm, one that could have been different, rather than just "business as usual".)

There is, however, a second paradigm, one that once held a much stronger foothold in the field but that has been largely stamped out, with remnants existing here and there like smoldering fires across a barren landscape. I call it *madness-as-strategy.* The core idea here is that when somebody is mad, everything inside of them is functioning exactly as it is supposed to, or as it ought, or as nature intended.

With this background set of beliefs, the purpose of the book—the job that it is meant to do—is not, in the first place, to demonstrate that this is an intellectually useful distinction. Rather, it is to dislodge the reigning worldview, *madness-as-dysfunction.* That's why I say the book could be read as a series of conceptual exercises designed to reveal *madness-as-dysfunction* as a distinctive and historically specific worldview, rather than the silent default of all theorizing. It is to lure *madness-as-dysfunction* out of its hiding place and expose it for what it is by contrasting it with an equally plausible but opposing alternative: *madness-as-strategy*. Crucially, as I emphasize in the book

> The question is *not* one of destroying madness-as-dysfunction or refusing to apply it where it deserves to be applied, but *to make it possible to even raise the question* of whether madness-as-dysfunction ought to be applied in any particular case. (Garson 2022, 2)

I'll close this synopsis by briefly outlining four main benefits of the book: for history, philosophy, Mad Pride, and mental health care.

First, the book provides a new set of tools to the historian to help them create new narratives about the history of madness—narratives that go beyond viewing this history merely as a clash between biological and psychological points of view, or even as a pendulum that swings back and forth between them. Incidentally, I do not accept what is sometimes described as the "biopsychosocial" point of view because it remains firmly within a certain controlling narrative that I would like to challenge, namely, that psychiatry's history is best viewed as a clash between biological and psychological perspectives. I return to this theme in my response to Khalidi.

Second, the book provides a new set of tools to the philosopher. *Madness-as-dysfunction* has become so entrenched in our current mental health landscape that it is often difficult to see *as* a distinctive worldview, one consolidated for fairly arbitrary (political, social, economic) reasons. One mark of its entrenchment is that philosophical definitions of the very concept of mental disorder or mental illness often rely on the concept of dysfunction, as if it were a necessary condition of what it is to be mad, as if the concept of dysfunction is "analytically contained" in the concept of mental illness. Wakefield (1992), for example, famously defined "mental disorder" in terms of harmful dysfunction, and Boorse (1977) defined "disease", whether mental or physical, in terms of the reduction of functional ability below typical efficiency.

The following point is crucial: I do not reject the Wakefieldian definition of mental disorder. I think Wakefield gives us the clearest and most philosophically defensible articulation of our current widely accepted notion of mental disorder. Moreover, I tend to agree with him that the content of the concept of mental disorder, as it exists today, is, roughly, the following: *a harmful psychological condition caused by an inner dysfunction* (e.g., Wakefield et al. 2006). However, if it turns out, as I suspect it will, that many of the psychological conditions that we currently refer to as "mental disorders" are, in fact, functional rather than dysfunctional, then that would give us a good reason to stop calling them "mental disorders"—and perhaps, in time, to abandon the category altogether.

The third benefit is that the book provides intellectual scaffolding for the movement known as *Mad Pride* or mad advocacy. "Mad Pride" is modeled on other progressive social movements, such as Black Pride or Gay Pride. It seeks to take an identity—in this case, being mad—that has often been disparaged or denigrated, and present it in a more positive light, and as a potential seat of political solidarity. Clearly, a crucial step in getting Mad Pride off the ground is to reject models that depict madness as merely a disease—a "condition" for doctors to "cure". Simply put, it requires a way to intellectually displace dysfunction-centered framings from their centrality in the mental health landscape. *Madness-as-strategy* does just this.

The fourth benefit, which I now see as the most crucial, has to do with changing the landscape of mental health care. (I would say "treatment", so long as the word is not taken to imply that madness or its various expressions are diseases to be cured, but forms of life that often require thoughtful support or empowerment.) I am inclined to believe that the entrenchment of *madness-as-dysfunction* in our current mental health landscape does a profound disservice to people who suffer with very real and debilitating problems, or in some cases, to people who possess cognitive styles that are marginalized or misunderstood, such as ADHD, autism, or dyslexia, conditions often presented under the banner of neurodiversity.

Take, for example, depression. If we see depression, as I think we should, as functional rather than dysfunctional—as, say, the brain's evolved signal that something in the environment is not going well rather than a "chemical imbalance"—this perspective utterly transforms the question of treatment. Quite simply, it means we ought to spend more time listening to what it's trying to say, rather than bombarding it with powerful psychoactive medications. (Of course, there are nuanced questions about when and under what conditions drugs might be useful for navigating mental health challenges, questions that I address in detail in Garson forthcoming a).

There is only one major respect in which my point of view has shifted since I wrote the book. I once envisioned the book as a plea to psychiatry to become more inclusive and pluralistic, as if I were encouraging psychiatrists along the following lines: "There are dysfunction-centered framings of mental health problems, and there are function-centered framings, and it would be good to have an open mind about this kind of thing. Perhaps some kinds of mental health problems, for some people, are best seen as dysfunctions, and others, as functions". But I have come to believe that psychiatry, in its inmost essence, is wedded to a dysfunction-centered framework (Garson forthcoming b). After all, psychiatry is a branch of medicine, and the goal of medicine is to treat or prevent diseases. To abandon a dysfunction-centered framing, then, is to cease practicing psychiatry. That is why I now think that mental health problems that are functional rather than dysfunctional do not in fact fall under the jurisdiction of psychiatry. For those, we must go *beyond* psychiatry. But that is a topic for another book.

REFERENCES

Boorse, Christopher. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44(4): 542–573.

Garson, Justin. 2022. *Madness: A Philosophical Exploration*. New York: Oxford.

Garson, Justin. Forthcoming a. "What Do Drugs Do? Rethinking Psychiatric Medications Outside of Biomedical Narratives." *Journal of Humanistic Psychology*.

Garson, Justin. Forthcoming b. "Beyond Psychiatry: Rethinking Madness Outside Medicine." In *Madness and Mental Health*, edited by Edward Harcourt. Cambridge: Cambridge University Press.

Wakefield, Jerome C. 1992. "The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values." *American Psychologist* 47: 373–388.

Wakefield, Jerome C. et al. 2006. "The Lay Concept of Conduct Disorder: Do Nonprofessionals Use Syndromal Symptoms or Internal Dysfunction to Distinguish Disorder from Delinquency?" *Canadian Journal of Psychiatry* 51: 33-39.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 2 | 101-115
https://doi.org/10.31820/ejap.21.2.1

MADNESS BY DESIGN: A GENEALOGY OF AN "ANTI-TRADITION"

Muhammad Ali Khalidi[1]

[1] City University of New York, USA

Discussion – Received: 29/09/2023 Accepted: 14/06/2024

This paper is part of a book symposium on Justin Garson's *Madness: A Philosophical Exploration* curated and edited by Elisabetta Lalumera (University of Bologna) and Marko Jurjako (University of Rijeka)

ABSTRACT

Psychiatric conditions are commonly regarded as mental disorders or dysfunctions of the mind. Yet there is a wealth of historical theorizing about the mind that conceives of these conditions as, in some sense, a matter of design rather than dysfunction. This intellectual legacy is the topic of Justin Garson's penetrating study, *Madness: A Philosophical Exploration* (2022). In this paper, I interpret Garson's book as a genealogy (in the Foucauldian sense) of the "anti-tradition" that he labels "madness-as-design". I argue that viewing the intellectual legacy that Garson analyzes through this genealogical lens has two benefits. First, it encourages us to identify other instances of madness-as-design (or madness-by-design), particularly those with an overtly political dimension, such as psychiatric conditions in a colonial context. Second, it should lead us to question the category of madness itself, which turns out to be radically disjointed, particularly since it cannot be unified under the rubric of disorder or dysfunction.

Keywords: psychiatry; mental disorder; dysfunction; genealogy; colonialism.

Correspondence: makhalidi@gc.cuny.edu

Introduction

There is a widespread contemporary assumption that psychiatric conditions can be uniformly regarded as mental disorders or dysfunctions of the mind. Yet there is a wealth of historical theorizing about the mind that conceives of these conditions as functional, useful, or adaptive, in other words, in some sense, as a matter of *design*. This rich intellectual legacy is the topic of Justin Garson's penetrating study, *Madness: A Philosophical Exploration* (2022). In this paper, I propose to read Garson's book as a genealogy of the "anti-tradition" that he labels, variously, as "madness-as-design" or "madness-as-strategy". Viewing the theoretical approaches that he analyzes through this genealogical lens has two benefits. First, it encourages us to identify other instances of madness-as-design (or madness-*by*-design[1]), particularly those with an overtly political dimension on the contemporary scene. Second, it should lead us to question the category of *madness* itself, which turns out to be radically disjointed, particularly since it cannot be unified under the rubric of disorder or dysfunction.

1. Madness-by-design as genealogy

Garson states at the outset that his work is neither a history of science, nor a genealogy. But I think it would be too hasty to take this overly modest statement at face value. The book certainly traces a particular theme through a number of historical authors and texts discussing various conditions of the mind. It is thereby a selective history of the topic that pays close attention to certain texts and their arguments, albeit less attention to their social contexts and political backgrounds. As such, it is more of an inquiry in the vein of history of philosophy than intellectual history,[2] meticulously reading the texts in question, from Hippocrates to Wakefield, and advancing interpretations of their views, without attempting to situate them in their respective historical eras in any detail. But the book can also be usefully seen as a genealogy, roughly in the Nietzschean or Foucauldian senses of the term.[3] If we take "madness" as the central concept under examination, there is arguably an attempt in this work to trace a neglected lineage for this category, and one that conforms

[1] Garson seems to use madness-*as*-design and madness-*by*-design interchangeably, and I will follow his lead in what follows.

[2] To use a distinction made by Schneewind, Skinner, and Rorty in the introduction to their edited volume, *Philosophy in History* (1984).

[3] Queloz (2021) provides an illuminating recent overview of the methodological tradition of genealogy, encompassing a range of thinkers from across the analytic-continental divide. In what follows, I will content myself with referring briefly to Foucault's characterization of genealogies to make a preliminary case that Garson's work on madness bears certain marks of a genealogy.

to some of the features that have been associated with the concept of genealogy, as I will try to argue in this section.

Garson has done philosophers of psychiatry (among others) a major service in excavating a history of theorizing about the mind that does not see mental conditions as disorders. Rather than mental disorder or dysfunction, this body of work regards "madness" as a matter of design or adaptation. An alternative title for this book might have been "madness by design", which is a phrase used by Garson at certain points to capture in dramatic fashion the radical idea that is being argued for here. Moreover, he groups many things under this rubric, including divine providence, natural law, physiological mechanisms, natural selection, and even malingering on the part of patients.[4] He shows how various historical figures have regarded allegedly dysfunctional mental conditions as products of natural or supernatural factors that are integral to the way that the world is ordered. Far from being aberrations or exceptions to the order of things, they are best regarded as a result of one or more of these factors, no less than the minds of "sane" or "normal" individuals. Garson characterizes the mad-by-design approach as an alternative to the standard view that madness is a matter of disorder or dysfunction:

> This alternative way of seeing proceeds from the conviction that some people are truly mad by design, that at least some of its forms are strategies for solving problems, coping with aspects of the environment, regulating one's mental economy. (Garson 2022, 10)

In doing so, he draws our attention to a neglected perspective in theorizing about the mind. Indeed, it would be more accurate to say that he elevates or promotes it to the level of a perspective, since it does not self-identify as such and it includes such variegated views. At the same time, Garson is at pains to emphasize that he does not regard "madness-as-design" as a tradition, even though it serves as the unifying theme of the book. In fact, he contends that its opposite, madness as dysfunction, *is* a tradition that "organizes itself into a proper narrative", but madness as design is "is like a child that interrupts that narrative from time to time, but each time in a different costume" (26). He also refers to madness-as-design as an "anti-tradition" (248).

[4] The phrase "madness by design" is used interchangeably with "madness as strategy" to denote the same basic idea. But even though the "strategy" locution is used more often to characterize the main thesis, to my ear, "madness by design" seems a more accurate characterization than "madness as strategy" for the body of work that he analyzes. For example, all but the last of the causal factors mentioned here would seem more aptly labelled a matter of "design" than "strategy."

However, I would argue that rather than characterize madness-as-design negatively, by contrast with its opposite, as a *non*- or *anti*-tradition, it might be cast in more positive terms. It is arguably not a tradition in the sense of a connected chain of ideas that refer to one another and build consciously on one another, but that does not prevent it from conforming to some of the features of a *genealogy*, as I understand them. For some of the writers Garson discusses, madness is a matter of divine retribution (25), while for others, it constitutes a means of redemption furnished by God (42), and for yet others it is a natural consequence of the way that our bodies are constituted (65). As he shows, there are a variety of ways in which this core idea is defended by various authors, some religious and others secular, some mentalistic and others biological, some evolutionary and others psychoanalytic. This chimes with Foucault's insistence that genealogy does not impose unity or fabricate an identity, but rather "fragments what was thought unified" (1977, 82).

Second, the historical lineage contains within itself many discontinuities rather than an uninterrupted course of continuous descent. Garson ingeniously finds evidence of the thesis even in authors who think of madness primarily as dysfunction. This is perhaps most prominent in the eighteenth and nineteenth century authors he discusses:

> (…) many of the theorists of this era, including Kant, Haslam, Wigan, Heinroth, Pinel, and even Griesinger, repeatedly uncover design inside of madness. It is as if, despite their best efforts to stamp it out, teleology persists; it cannot be entirely canceled or negated. (Garson 2022, 76)

Effectively reading some of these texts against themselves, Garson shows how even Kant, who viewed madness as a series of disorders of our mental faculties, endorses the idea that madness has a goal-directed character, since "defective reason still attempts to systematize its mad productions" (90). Similarly, Wigan, who conceives of madness as the product of a divided brain, holds that "Madness must harness or co-opt reason; it must compel reason to serve its own perverse end", much as autoimmune diseases harness the body's immune system to a harmful end (111). In these cases and others, we see "accidents", "minute deviations", and "complete reversals" in the history of the notion of madness-by-design, just as was posited by Foucault (1977, 81) for genealogy in general.

Finally, Garson resists the temptation to distill an essence out of this tendency or to anchor it in a single origin. Despite the fact that he begins his story with ancient conjurors and magicians, and notwithstanding the fact that he sometimes detects echoes of this ancient line of thought in later authors, he is also at pains to deny that the authors under discussion can be regarded as heirs to the ancient magicians. The concept of madness-as-design does not have a single origin or ancestor and it consists of "resonances" (26) and echoes rather than descendants or offspring. This, again, squares with Foucault's disavowal of the quest for origins and insistence on "numberless beginnings" (1977, 81).

Even though there is not a single tradition of madness-as-design, Garson's narrative is not just a selective history of psychiatry, foregrounding a number of eccentric contributions. The features I have highlighted, including multiplicity, discontinuity, and absence of origins or foundations, seem to be recurrent properties of genealogies, as posited by writers like Foucault and others. Moreover, I would argue that seeing madness-by-design as a genealogy has two benefits. First, it should encourage us to look for other exemplars of theorizing about madness in this vein, and hence, to reconceive psychiatric conditions without the burden of dysfunctional thinking or the concept of disorder. In particular, I would argue that it highlights the value of understanding at least some psychiatric conditions as expected responses to certain social and political regimes. Second, it should lead us to question not just the concept of "mental disorder"—since the second half of that label is now rendered moot—but indeed even a unified notion of psychiatric condition, or the idea that there is such a thing as "madness" in the first place. In the rest of this paper, I will pursue both of these leads. In section 2, inspired by Garson's genealogy, I will make a suggestion as to how one could extend his account to other significant cases of madness-by-design, notably those that have an overtly political character. In section 3, I will try to use Garson's analysis to push back against the notion that there is such a thing as madness in the first place, that is to say, against the idea that madness is a real or natural kind.

2. Madness as a matter of social and political design

Garson has done a prodigious amount of research on a fascinating cast of characters in the history of psychiatry. While some are well known (e.g. Burton, Krapelin, Freud), others are much less so (e.g. Wigan, Griesinger, Goldstein), and yet others are well known but not for psychiatry (e.g. Locke, Kant). These figures represent a range of different approaches to conceiving of madness in terms of design, but one aspect of the madness-by-design perspective that does not appear to be in evidence before the twentieth century has to do with the social and political causes that can give rise to madness.[5] Greater awareness of the social factors in the ontogenesis of madness may be one virtue of late twentieth and early twenty-first century psychiatric theorizing, as the "biopsychosocial" model of madness has gained ascendancy in the past few decades (Bolton and Gillett 2019).[6] However, the *political* element is seldom highlighted as a separate causal factor, giving rise to what might be labelled a "biopsychosocio-political" model. In Garson's genealogy, the most overtly political spin on madness-as-design can be found in the work of R. D. Laing in the mid- to late-twentieth century, but Laing's analysis of the relationship of schizophrenia to capitalism is less than convincing and not widely credited, at least nowadays. Nevertheless, the political dimensions of madness are worth taking more seriously and can be illustrated by the analysis of other forms of madness provided by other theorists, or so I will try to argue in this section.

An instructive instance of political madness-by-design may be found in the recent colonial past. As is well known, the great theorist of colonialism, Frantz Fanon, was trained as a psychiatrist and dedicated a large section of his work, *The Wretched of the Earth* (1963), to a series of psychiatric case studies meant to illustrate the ways in which the realities of colonialism, military occupation, indiscriminate killing, torture, rape, house demolitions, and a host of other forms of violence inflicted on the natives by the settlers impacted their mental health. Fanon even refers to colonialism "in its essence" as "taking on the aspect of a fertile purveyor for psychiatric hospitals" (1963, 249). While disavowing writing a scientific work and eschewing arguments over "nosology" and "therapeutics" (1963, 251), Fanon posits that colonialism, "is a systematic negation of the other person and a furious determination to deny the other person all attributes of humanity" (1963, 250).

Though Fanon does not attempt to provide a causal or mechanistic account of the ways in which denial of humanity might lead to depression, delusion, impotence, homicidal compulsion, and a myriad other mental conditions, he evidently recognizes these symptoms of madness as predictable outcomes of the pathological state that is colonialism. He states that some of his cases are clearly "reactionary", being the direct effects of the unspeakable crimes of colonialism, but most "give evidence of a much more widely spread causality although we cannot really speak of one particular event giving rise to the disorders" (1963, 252).

[5] One possible exception in Garson's narrative occurs in the work of Haslam (1764-1844), who was "adamant that social, psychological, and environmental factors can trigger the inner pathology that generates madness" (97).

[6] For recent discussion of the biopsychosocial model, see the papers in the special issue of EuJAP, guest edited by Cristina Amoretti and Elisabetta Lalumera (2021).

To illustrate, he relates a case of homicidal impulses on the part of an Algerian who was the survivor of a mass murder in his village by the French authorities (1963, 259-61). He also describes a case of depression, hallucinations, and "anxiety psychosis" in an Algerian man whose mother was killed by the French and who had subsequently killed an unarmed woman colonist while fighting with the resistance (1963, 261-4). He details cases of "noise phobia", insomnia, and sadistic tendencies in a group of children whose parents had been killed by the French and who had been displaced by fighting and sent to live in Morocco or Tunisia (1963, 277-8). Fanon dedicates a separate section to the symptoms of victims of torture, which include standard psychiatric conditions such as depression and apathy, as well as more recherché symptoms such as "electricity phobia" (specifically in victims of torture by electricity), inhibition, and phobia "of all private conversations" (1963, 280-293). Moreover, he thinks that colonialism does not just produce madness in its primary victims, colonized peoples, but in the colonizers and the enforcers of the colonial order, including the French police inspector who tortures his wife and children (1963, 267), or the French policeman who develops depression as a result of his participation in the torture of Algerians (1963, 264).

Perhaps the most disturbing case discussed by Fanon is that of two Algerian boys, ages 13 and 14, who kill their French playmate on the grounds that "the Europeans want to kill all the Arabs" (1963, 271). They go on to explain: "We can't kill big people. But we could kill ones like him, because he was the same age as us" (1963, 271). When asked why they chose to pick on their friend in particular, their matter-of-fact response is: "Because he used to play with us. Another boy wouldn't have gone up the hill with us" (1963, 271). The childish directness of their answers combined with the cold-blooded ruthlessness of their reasoning are related without comment by Fanon, notably without attempting to impute causality. But it is clear that he thought that these and other behaviors that he saw in his clinical practice were an integral feature of colonialism.

After detailing the ways in which French colonial psychiatrists attributed to Algerians in particular, and North Africans or Arabs in general, a criminal mentality and character traits of aggressivity, lack of emotivity, persistent obstinacy, mental puerility, and impulsivity, among others (1963, 294-304), Fanon proceeds to offer an alternative account:

> The Algerian's criminality, his impulsivity, and the violence of his murders are therefore not the consequence of the organization of his nervous system or of characterial originality, but the direct product of the colonial situation. (Fanon 1963, 309)

In other words, it is the anticipated effect of a situation of systemic violence and dehumanization. For Fanon, the pathologies of his patients are the normal response to a pathological system of oppression. Indeed, he turns the tables on colonial psychiatry by asserting that the alleged laziness and intransigence of natives under colonial domination are in fact not pathologies at all, but the natural state of resistance to colonialism. He writes:

> How many times—in Paris, in Aix, in Algiers, or in Basse-Terre—have we not heard men from the colonized countries violently protesting against the pretended laziness of the black man, of the Algerian, and of the Viet-Namese? And yet is it not the simple truth that under the colonial regime a *fellah* [Arabic for farmer or peasant] who is keen on his work or a Negro who refuses to rest are nothing but pathological cases? The native's laziness is the conscious sabotage of the colonial machine (…). (1963, 294)

If there is dysfunction here, it is to be found not in the colonized people but in the political regime of colonialism, which is the real site of pathology. The colonial situation seems to fit well within Garson's general rubric and would make a valuable addition to the genealogy of madness-by-design, specifically one with an overtly political dimension.

This political extension of Garson's genealogy helps relocate some sources of pathology in the contemporary world, by displacing them from the individual to the broader political context. It would be a mistake to think that the colonial era is entirely a thing of the past, since colonialism persists in the world in such places as Palestine/Israel, where a century of colonization has wreaked havoc on mental health. Here, too, colonial dispossession and denial of self-determination can be seen as a "fertile purveyor for psychiatric hospitals" (though hospitals, clinics, and trained professionals are exceedingly scarce in occupied Palestine; see Giacaman et al. 2011). Moreover, this political setting also reveals the inadequacy of standard psychiatric categories such as "post-traumatic stress disorder" (PTSD), which ignore the political context or implicitly assume a default political context from the global North. Although PTSD is the most commonly reported psychiatric condition among Palestinians in the occupied territories, Palestinian psychiatrist Samah Jabr articulates the problematic nature of the diagnosis:

> In Palestine, traumatic threats are ongoing and enduring. There is no "post-traumatic" safety. The phenomena of avoidance and hyper-vigilance are considered to be dysfunctional psychological

> reactions in a soldier who has returned to the safety of his hometown. But for tortured Palestinian prisoners, such symptoms are reasonable reactions, insofar as the threat lives on; they may be re-arrested and tortured again at any time. (Jabr 2019)[7]

In such contexts, psychiatric conditions are not rightly seen as disorders at all but as indicators of a disordered political regime:

> It is therefore essential to focus on the effects of the Israeli occupation on the mental health of the Palestinian people and to advocate for their national and human rights. Otherwise, the experiences of Palestinians will be pathologized and their responses medicalized while the status quo of the pathogenic context remains the same. (Hammoudeh et al. 2020, 84)

Other researchers on mental health and well-being have also warned that ignoring the "driving force of political conditions" risks locating the source of pathology in individuals rather than political contexts (Barber et al. 2014, 101).

The enduring relevance of madness-by-design as a political phenomenon can also be demonstrated with reference to its recurrence in the era of global climate change. In the Anthropocene, the mental condition of what has been called "climate anxiety" (also "climate panic" or "eco-anxiety") has become widespread, and is arguably a rational, or at least natural, response to the climate emergency. As one researcher puts it, again displacing pathology from individuals to their circumstances: "the climate crisis does not just *induce* trauma under certain circumstances—it is a new form of trauma that pervades the circumstances of our life" (Woodbury 2019, 1; original emphasis).

Some writers have distinguished two ways in which climate change impacts mental health. The first includes the direct influence of extreme weather events and natural disasters on people's states of mind, for example those who have been displaced or forced to migrate as a result of climate change. The second involves anxiety about climate change, which can affect even those who have not experienced direct impacts and can include concerns about harm to future generations. Particularly when it comes to the latter, Clayton stresses that "[i]t is important to avoid pathologizing the emotional response to climate change" (2020, 3). As in

[7] She also notes that the category "fails to capture the experiences of communities living with collective historical trauma" (Jabr 2019).

the case of colonialism, a pathologizing approach can serve to direct "attention toward individuals and away from the social causes and possible social responses to climate change" (Clayton 2020, 3). Moreover, intervention techniques that rely on cognitive reframing to de-emphasize or deny the threat, are both unlikely to be effective and do not promote the well-being of society at large (Clayton 2020, 4). Indeed, they would seem to be morally and politically reprehensible. By contrast, there is some evidence to suggest that working to mitigate climate change is a more effective intervention and some studies show positive correlations between happiness and "pro-environmental or sustainable behavior", based on research conducted in Mexico (Corral-Verdugo et. al 2011, 102). Far from being futile, political resistance may improve mental health as well as change policy. This means that supposedly extreme reactions to the climate emergency cannot be seen as dysfunctional or disordered, but in some sense at least, as a matter of design.

3. Is there such a thing as "madness"?

As already emphasized, there is a great deal of variety in the ways in which forms of madness can be seen as instances of design, as opposed to disorder or dysfunction. Garson's philosophical history of madness-by-design presents us with numerous routes to conceiving of mental conditions as "features" rather than "bugs", ranging from punishments for sins to biological adaptations. Indeed, for some of the thinkers discussed by Garson, it would be a stretch to say that madness is a matter of design from the perspective of the authors themselves. At best, it emerges from his inventive interpretations of their work; indeed, in some cases, it requires a kind of "contrapuntal reading" of the texts (cf. Said 1993). This is not a problem for a genealogical account, which is meant to be disunified and disjointed, but it may be a problem for any attempt to delineate a unitary concept of *psychiatric disorder* (or closely related concepts, such as *psychiatric condition*, *mental disorder* or *mental illness*).

In this section, I will argue that reflection on madness-by-design ultimately serves to undermine the existence of a unified category of *psychiatric disorder* and related categories. This conclusion has been pressed by others, but I think that Garson's inquiry gives us further reason to doubt the validity of such a construct. In making this case, I will adopt a realist but non-reductionist approach, according to which scientific categories (including psychiatric ones) aim to identify natural (or real) kinds, and that these kinds are associated with aspects of the causal structure of the world (including the structure of the human mind).[8]

Many attempts to characterize psychiatric conditions consider them to be dysfunctions and regard dysfunction as a necessary condition for something to be a psychiatric condition. But if the madness-by-design perspective is an apt characterization of at least some psychiatric conditions, then there does not appear to be a common denominator among all these conditions. While some may result from biological dysfunctions pertaining primarily to the individual and would recur across a broad range of social environments, others may just be functional responses to social stressors, for example. Given the heterogeneity of their central features as well as their causes and effects, there would seem to be no basis to group them together as members of a single kind.

Even if one disagrees with many of the specific claims made by proponents of the madness-by-design perspective, collectively they lend credence to the idea that there is a great deal of heterogeneity among the conditions lumped together in the category of psychiatric disorder. This heterogeneity is not alleviated if we substitute the label “psychiatric disorder” with such terms as “psychiatric condition”, “neurodiversity”, or even “madness”, since it pertains to the assortment of conditions that are generally grouped together under these labels. The category of *psychiatric disorder* is thought to include such diverse conditions as autism, depression, schizophrenia, and post-traumatic stress disorder, which seem to have nothing in common apart from the fact that they are conditions of the mind-brain that are commonly thought to be dysfunctional in some way. But if mind-brain dysfunction is *not* a common denominator, then it cannot give unity to the category, and there does not seem to be anything else that pertains to psychiatric conditions as such.

This claim about the category *psychiatric disorder* (or *madness*) does not prevent some *specific* psychiatric conditions from being real kinds (e.g. *schizophrenia*, *autism*) (cf. Beebee & Sabbarton-Leary 2010). Although that might seem paradoxical at first sight, there is no tension in principle between asserting that a superordinate category does not correspond to a real kind while some of its subordinate categories do. (Compare: *pet* is probably not a real kind either in biology or the social sciences, but *dog* and *goldfish* are real biological kinds.) Now, it may be thought that the same obstacles to kindhood that apply to the superordinate category also apply to the subordinate categories. But this does not seem warranted,

[8] I will not try to justify this background assumption here, but I have tried to defend it elsewhere; see e.g. Khalidi (2013, 2023).

since at least some psychiatric conditions are relatively homogeneous, unlike the superordinate category that they are usually subsumed under.

One might object to this proposal to eliminate the category of *psychiatric disorder* on the grounds that it would undermine the very basis of psychiatry. If the category *psychiatric disorder* does not correspond to a real kind and the conditions that it studies are disunified and not subsumed under a single umbrella, should the field splinter into a number of different disciplines, each dedicated to one or a subset of conditions? Would this not have an adverse effect on both empirical research and clinical practice? And will it lead eventually to the elimination of psychiatry as a viable branch of medicine?

The implications of this conclusion for psychiatry are significant but need not lead to such consequences. There are other domains of medicine that study diverse sets of phenomena. After all, pediatrics investigates and treats a variety of different conditions that affect children, though there is no unified category of “children’s disease” or “sick kids”. Some of these conditions are rightly regarded as dysfunctions while others are not, for example symptoms like fever, vomiting, and diarrhea that help the body to combat bacterial infection, or allergic reactions that serve to stave off allergens (cf. Lillienfeld and Marino 1999). Similarly, psychiatry can be conceived as a discipline that focuses on a range of conditions pertaining to the mind-brain, using a diverse set of methods and deploying a wide variety of interventions.[9] But rather than view all of these phenomena as disorders, it would be better to regard them as distinctive mental conditions or dispositions.

It is also worth considering another objection to denying that there is a valid scientific category that groups together all psychiatric conditions. It might be said that the above considerations apply to the category of *psychiatric disorder*, but not *madness*. However, the two terms are roughly coextensive and both are used to denote a heterogeneous collection of conditions, at least some of which may be adaptive in various contexts. Since that applies to the set of conditions discussed under this general rubric, whichever label we use, the substantive point is the same. But, it might be protested, if we eliminate *madness* as a category, that might have lamentable consequences, since especially when reclaimed by those who are collectively labelled as “mad”, it can result in solidarity and a sense of common cause, as in the movement for “mad pride”. To be clear, I have cast doubt on the concept primarily as a category for scientific research,

[9] Moreover, some of these conditions may require social or political interventions rather than individual treatments, while others may not be amenable to or in need of treatment at all.

not one for building support and solidarity among those who have been pathologized. If it is understood as a category that aims not at identifying a real kind of condition, but one that can serve a moral or political purpose, then it may be worth retaining, as long as we do not consider it to identify a real kind in the biomedical or social sciences. Even though this may not have been Garson's intention, I submit that his project leads us inevitably to question the category *madness* itself. The category can be seen to have played the role of Wittgenstein's ladder in his inquiry: a useful implement for reaching a destination that can be dispensed with by the end of the exercise.

4. Conclusion

In this paper inspired by Garson's book *Madness: A Philosophical Exploration*, I have tried to make a case for three broad claims. The first is that despite his demurrals, Garson has effectively provided a genealogy (roughly in the Foucauldian sense) of madness-by-design, a historical perspective on psychiatric conditions that conceives of them as (in some sense) a product of design rather than dysfunction. The second is that this genealogy can be used to identify other instances of adaptive or designed psychiatric conditions in the contemporary world, notably those with an overtly political dimension. The third is that Garson's genealogy ultimately leads us to question the superordinate category *madness*, since this historical exploration lends further support to the contention that it is thoroughly heterogeneous. To put it more succinctly: there is no such thing as madness and *Madness* is a book about it.[10]

Acknowledgments

I am very grateful to Tarif Khalidi, Itsue Nakaya Perez, Diane Riskedahl, and Maya von Ziegesar for comments that led to numerous improvements, and to Justin Garson for illuminating conversations. Thanks also to the editors of this special issue for their suggestions.

REFERENCES

Amoretti, Maria Cristina, and Elisabetta Lalumera. 2021. "Introduction to the Book Symposium on The Biopsychosocial Model of Health

[10] Apologies to Steven Shapin, whose book, *The Scientific Revolution* (1996), opens with the sentence: "There was no such thing as the Scientific Revolution, and this is a book about it".

and Disease by Guest Editors." *European Journal of Analytic Philosophy* 17 (2): M1-8. https://hrcak.srce.hr/en/broj/20824

Barber, Brian K., Carolyn Spellings, Clea McNeely, Paul D. Page, Rita Giacaman, Cairo Arafat, Mahmoud Daher, Eyad El Sarraj, and Mohammed Abu Mallouh. 2014. "Politics Drives Human Functioning, Dignity, and Quality of Life." *Social Science and Medicine* 122: 90-102.

Beebee, Helen and Nigel Sabbarton-Leary. 2010. "Are Psychiatric Kinds 'Real'?" *European Journal of Analytic Philosophy* 6 (1): 11-27.

Bolton, Derek, and Grant Gillett. 2019. *The Biopsychosocial Model of Health and Disease: New Philosophical and Scientific Developments*. Cham: Palgrave Macmillan.

Corral-Verdugo, Victor, José F. Mireles-Acosta, Cesar Tapia-Fonllem, and Blanca Fraijo-Sing. 2011. "Happiness as Correlate of Sustainable Behavior: A Study of Pro-ecological, Frugal, Equitable and Altruistic Actions that Promote Subjective Wellbeing." *Human Ecology Review* 18 (2): 95-104.

Clayton, Susan. 2020. "Climate Anxiety: Psychological Responses to Climate Change." *Journal of Anxiety Disorders* 74: 102263.

Fanon, Frantz. 1963. *The Wretched of the Earth*. New York: Grove Press (first published in French in 1961).

Foucault, Michel. 1977. "Nietzsche, Genealogy, History." In *The Foucault Reader*, edited by Paul Rabinow, 76-100. New York: Pantheon Books.

Garson, Justin. 2022. *Madness: A Philosophical Exploration*. Oxford: Oxford University Press.

Giacaman, Rita, Yoke Rabaia, Viet Nguyen-Gillham, Rajaie Batniji, Raija-Leena Punamäki, and Derek Summerfield. 2011. "Mental Health, Social Distress and Political Oppression: The Case of the Occupied Palestinian Territory." *Global Public Health* 6 (5): 547-559.

Hammoudeh, Weeam, Samah Jabr, Maria Helbich, and Cindy Sousa. 2020. "On Mental Health amid COVID-19." *Journal of Palestine Studies* 49 (4): 77-90.

Jabr, Samah (2019). "What Palestinians Experience Goes Beyond the PTSD Label." *Middle East Eye*, 7 February 2019 https://www.middleeasteye.net/opinion/what-palestinians-experience-goes-beyond-ptsd-label.

Khalidi, Muhammad Ali. 2013. *Natural Categories and Human Kinds*. Cambridge: Cambridge University Press.

Khalidi, Muhammad Ali. 2023. *Natural Kinds*. Cambridge: Cambridge University Press.

Lillienfeld, Scott O. and Lori Marino. 1999. "Essentialism Revisited: Evolutionary Theory and the Concept of Mental Disorder." *Journal of Abnormal Psychology* 108 (3): 400-411.

Queloz, Matthieu. 2021. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*. Oxford: Oxford University Press.

Said, Edward W. 1993. *Culture and Imperialism*. New York: Vintage.

Schneewind, Jerome B., Quentin Skinner, and Richard Rorty. 1984. *Philosophy in History: Essays in the Historiography of Philosophy*. Cambridge: Cambridge University Press.

Shapin, Steven. 1996. *The Scientific Revolution*. Chicago: University of Chicago Press.

Woodbury, Zhiwa. 2019. "Climate Trauma: Toward a New Taxonomy of Trauma." *Ecopsychology* 11 (1): 1-8.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 2 | 117-132
https://doi.org/10.31820/ejap.21.2.2

STRATEGY, PYRRHONIAN SCEPTICISM AND THE ALLURE OF MADNESS

Sofia Jeppsson[1] and Paul Lodge[2]

[1] Umeå University, Sweden
[2] University of Oxford, United Kingdom

Discussion – Received: 04/07/2023 Accepted: 01/10/2023

This paper is part of a book symposium on Justin Garson's *Madness: A Philosophical Exploration* curated and edited by Elisabetta Lalumera (University of Bologna) and Marko Jurjako (University of Rijeka)

ABSTRACT

Justin Garson introduces the distinction between two views on Madness we encounter again and again throughout history: Madness as dysfunction, and Madness as strategy. On the latter view, Madness serves some purpose for the person experiencing it, even if it's simultaneously harmful. The strategy view makes intelligible why Madness often holds a certain allure—even when it's prima facie terrifying. Moreover, if Madness is a strategy in Garson's metaphorical sense—if it serves a purpose—it makes sense to use consciously chosen strategies for living with Madness that don't necessarily aim to annihilate or repress it as far as possible. In this paper, we use our own respective stories as case studies. We have both struggled to resist the allure of Madness, and both ended up embracing a kind of Pyrrhonian scepticism about reality instead of clinging to sane reality.

Keywords: madness; Pyrrhonian scepticism; mania; psychosis; psychiatry.

Correspondence: sofia.jeppsson@umu.se

Introduction: Madness as strategy

In Justin Garson's groundbreaking *Madness: A philosophical exploration* (2022), he takes us on a tour through history and the constantly resurfacing tension between seeing Madness as a *dysfunction* and seeing it as a *strategy*. On the dysfunction view, Madness is analogous to a physical problem like asthma—when the asthmatic's bronchi close up and he can't get enough air, the breathing apparatus is dysfunctional, not working the way it should. On the strategy view, Madness is analogous to a physical phenomenon like fever—when you're infected with a pathogen and your body temperature rises, this helps your body to heal by slowing down pathogenic reproduction (ibid., 90). As the analogy shows (and *pace* Kraepelin, ibid. 78), "strategy" doesn't imply "consciously chosen", merely that it serves a purpose.

Garson points out that if we view Madness as a strategy, as something which may fill a function for the Mad person, this may have implications for treatment decisions. Fever, though purposeful, may become harmful in itself if it goes too high, and Madness may harm as well. Nevertheless, if we see it as strategic, we're less likely to try to repress all Mad phenomena at all costs, more likely to look beneath "symptoms" to see what they might be a response to (ibid., 10-11). Moreover, insofar as Madness provides a way of dealing with problems in your life or allows you to *escape* from said problems (ibid., 125, 128, 130-131, 174), it can be understandably *tempting*. Madness is normally spoken of as an affliction that befalls people, but it might also be something that draws you in.

Garson cites Arthur Wigan, who talks of how "the sick brain" tries to seduce "the healthy brain" into Madness, by presenting the person with a tempting alternative worldview in which he's, e.g., a great and important leader or in touch with God himself, instead of a hallucinating madman (ibid., 127-131). Descriptions of Madness as alluring and seductive is far less common in modern times, but Edward M. Podvoll's *The Seduction of Madness* (1991) stands out as an exception. Like Wigan, Podvoll talks about how Madness can be tempting and draw people in. Moreover, both believe that recovery must entail a wholesale rejection of Madness in favour of reason. But this is not the only possible solution for a person struggling with Madness. Madness can be embraced as well, and given a more positive spin, as in the Mad Pride movement (Garson 2022, 12).

We, Paul Lodge and Sofia Jeppsson, can both relate to the view of Madness as meaningful, strategic and alluring. Our respective Mad experiences left us both struggling with philosophical problems about what to believe, what to do, and whether to cling to sanity or go fully Mad again.

Moreover, we find ourselves having settled on a somewhat similar way of managing the tension that arises when these forces are at play. We believe that our respective stories provide interesting illustrations of how seeing Madness as a simple brain dysfunction which should be fixed can be of little help, and even profoundly unhelpful, and offer an alternative framework for coping with Mad existence. However, we would also like to stress at the outset that we regard this as one strategy—in the consciously chosen sense of the word, not in the fever-analogy sense—that has in fact worked for us. We hope it might work for others, but it also seems clear that many other strategies may be required.

Thus, we frequently use “strategy” in a different sense than Garson’s. We do not aim in this paper to take a view on whether the *combination* of Madness and our conscious strategies for dealing with it should be understood as a strategy in the Garson sense, as a dysfunction, or indeed in some other way. However, some of the comments that we make do engage with this question, which is clearly worthy of further consideration and one that we hope to address in future work.

Paul’s story

I received a bipolar diagnosis in 1994 when I was twenty-six and studying for my PhD in New Jersey. It was at that point that I had my one and only manic episode. Whilst the formal diagnosis did not occur until my mid-twenties, there were clearly signs much earlier. For the last two years of high school, I suffered from what I now take to have been a significant period of major depression and I have vague recollections of depressive phases and strange ‘quasi-mystical’ experiences earlier in life.

Bipolar disorder is so-named because most people with the label have experienced periods of both depression and mania. However, it is the latter that I have in mind when I speak of myself as having experienced Madness. The formal criteria on the basis of which people receive a diagnosis of a manic episode are usually those found in the Diagnostic and Statistical Manual of Mental Disorders, though only some of those criteria will be relevant for the current article: in particular, being “more talkative than usual” having “flights of ideas” and a sense that one’s “thoughts are racing”, “distractibility”, and “inflated self-esteem or grandiosity” where these give rise to “marked impairment in social or occupational functioning” (DSM-5, 124).

Given that these criteria are diagnostic they focus for the most part on aspects of the manic subject which are observable by clinicians or readily reported to clinicians by the subject. Thus, they do not attempt to speak at length to what it is like to be undergoing such an experience. And to this extent, they do not point toward all of the challenges my manic episode has posed for me. In particular, they do not offer any purchase on why mania has an allure. However, I think it is possible to enrich the account of the experience in ways that do speak to this.

Two crucial things are missing from the criteria listed above. The first is something that unifies them all; namely that they are aspects of a way of responding to a disruption in what it is like to be. I think it is possible to convey at least some sense of what this unifying feature is. The use of the term “inflated” in connection with the sense of self offers a clue. It points to the way in which manic subjectivity expands as the sense of there being exponentially more and more to attend to breaks into consciousness. Moreover, I think this sense of there being more to attend to allows us to make sense of why the manic subject is distractible, has flights of ideas, and racing thoughts. For this can be understood as the mind responding to the increase in what is present to it by relying on its already developed capacities to conceptualize things. It is important that I used the term “exponentially” above. Indeed, another way to articulate my sense of what was happening as mania took hold was that I was being overwhelmed by a rapidly increasing amount of reality and trying my best to comprehend that by using concepts I already had to forge non-standard links. Another aspect of this experience that is alluded to in DSM-5 was the grandiosity that accompanied this. And this is perhaps unsurprising. During my period of manic subjectivity, I took myself to be seeing more of what there is than anyone else had ever seen and gaining greater insight; and, as is often the case for manic subjects, this apparent insight seemed so profound that it came to express itself via a sense of an almost messianic destiny to reveal the truth about existence to others.

The second crucial thing that is missing from the DSM criteria concerns how it feels to be in the grip of mania. DSM-5 talks of “elevated, expansive, or irritable mood”. Whilst these terms capture something, they fail to do justice to the way in which some phases of mania are intoxicating. There was irritability at times—mainly in the presence of others who were not able to see what I was seeing. But for the most part I was overflowing with an ecstatic joy which attended a sense that I was experiencing the way in which reality was showing itself to be more full of meaning, and with a sense of gaining limitless access to the perfection of both the subject and object of the experience.

So far, the ways in which I have talked about mania might suggest that it can be helpfully characterised as strategic in Garson's sense. But, as noted above, our concerns in this paper are with strategies of a different kind, namely the strategies we have adopted for living with our respective Madness. In my case, this is itself something of a bipolar issue. On the one hand my recollection of my manic episode is very negative. In particular, there are memories of the ways in which it seriously undermined my ability to maintain social relationships. But this is mixed up with the recollections of being caught up in something ecstatically revelatory, a recollection which was, for a long time, combined with a sense of there being unfinished business to attend to. Unsurprisingly, it is the latter which has been the source of the allure that my Madness held prior to the adoption of the sceptical strategy that we will discuss below.

In light of the socially destructive aspects of my mania, one of the things that I have done ever since my episode is take drugs under the supervision of psychiatrists. Initially, this was forced upon me; but soon after it became voluntary. My compliance speaks to an overriding desire never again to experience the alienation that I associate with having my manic episode. The drugs worked—and still work—well enough, but for a long time, I felt frustrated taking them. I regarded medication as a regrettable trade-off. Living with others was prioritized over making further sense of the mania and the things that seemed to have been ecstatically revealed. For it also seemed to me that making further sense would only be possible by stopping my regimen of drugs and becoming manic again.

For all that I myself lacked a strategy to address this pull to make further sense of my mania, it was also clear to me that others who have had manic experiences do. We can usefully think of these as inflationary or deflationary: inflationary insofar as they involve taking the having of further manic experiences to be valuable, and deflationary insofar as they do not. However, I am interested here only in those which are deflationary, given that the approach on which I have finally settled at this point is of this kind.

One deflationary strategy is built into the way in which some people rely on drugs in the wake of mania. Here I am thinking of people who conceptualize their sense of themselves and the world as metaphysically dependent on the brain and its properties. I will refer to this as "materialism" for convenience's sake. For the materialist, manic subjectivity, like any other, is dependent on the way in which the brain is functioning at a given point in time, with mania as a kind of *dys*function. It is a dysfunction, in part, because it leaves one unable to conceptualize things in the ordinary reality-revealing way. But luckily (for some at least) it can be combatted by the ingestion of drugs which alter the structure of the brain's chemistry so that normal functioning is regained. This perspective offers an additional advantage for some. Rendering mania intelligible in this way may also neutralize some of its allure. If manic episodes are conceived as due to changes in brain chemistry, there is perhaps less pressure to take seriously any tendency to regard them as revelatory. Whilst memories of such states may involve a sense of revelation and ecstatic affect, this is likely to be regarded as delusional; and, whilst there may still be some attraction to the affective component of mania, this is likely to be significantly reduced insofar as it is decoupled from the sense that the experience was revelatory.

I was already somewhat suspicious of reliance on this kind of account of our mental lives before my manic experience. But it has proved impossible for me to appeal to anything of this kind in its wake since materialism doesn't speak at all to my recollection of the changes in the sense of self that attended that the mania. I have always remembered the experience as involving something that simply isn't rendered intelligible as a manifestation of dysfunction in a material system. In occupying this position, I take myself to have been in a similar predicament to many other manic subjects for whom subsequent appeals to materialism and dysfunction seem inadequate. However, I have also been unable to avail myself of another deflationary strategy that some take at that point, namely those for whom the sense of revelation remains, but in such a way that it is amenable to an alternative metaphysical explanation.

Here I am thinking primarily of those for whom mania leads to a life which involves some kind of spiritual conversion. Such a subject might take themselves to have had an experience of divine presence, for example. But whatever the precise content, the common denominator with the kind of response that I have in mind is a response that regards the experience as an encounter with a reality the nature of which can be rendered intelligible to at least some degree, and that does not need to be repeated in order for its work to be done. Crucially important as the initial occurrence may have been, the revelatory significance of the manic experience can now be understood in such a way that there is no need to become manic again in order to reap the benefits of the revelatory significance. In such instances, it is also likely to be true that the experience can then be shared with others for whom similar interpretations seem appropriate. And the initial social estrangement brought on by the mania may turn out to be a gateway to the very opposite, namely membership of a community which is forged around taking the mania itself to be something that binds that community together. For better or worse, none of the alternative metaphysical explanations that I knew of prior to my mania or which I investigated in response to its occurrence helped me make sense of the experience. The difficulty I faced with materialism and the other metaphysical views was that my memory of the manic experience included a still compelling sense that I experienced reality in a way that outstripped all the available attempts to comprehend it.

For a long time, my fear of the social consequences of becoming manic remained a primary determining factor in my relationship to my manic experience. Perplexed and exhausted, I turned away from any attempt to engage directly with its significance and I took the drugs prescribed for me to try to stave off any recurrence. But I was unable to shake the allure of mania and the sense that I was denying myself something that I regarded as crucially important, namely the possibility of a revelation of a truth of great significance. Indeed, the temptation was to think that the denial of this was simply due to the constraints of social conventions that I would have rather had the courage to ignore.

However, over the past five years or so a change took place. I remained firm in my resolve to continue taking my medication, but it also seemed imperative that I find some way to engage fully with the allure of mania. Rather than living with a sense of fragmentation, I was drawn back by the desire to find the sense at the end of the manic rainbow. And, at this point, I found myself embracing a mode of being that in hindsight seems to have been trying to force itself on me all along, which I will call “Pyrrhonian scepticism”.

Like most terms of art, “Pyrrhonian scepticism” is explicated in different ways by different people. However, I use it to point toward a number of core components of my current existence. Central to this has been a reconceptualizing of the significance of the manic experience that I had. I no longer interpret the memory of my manic experience as the memory of a mode of being in which I had been gaining insight which was cut short. Rather it has come to seem more appropriate to think of it as an experience which could not have but been cut short. For what now seems to me to have been the case when I recall the experience is that it was one of trying to use my conceptual capacities to make sense of something which essentially outstripped those capacities. For want of a better expression, it seems to have been an experience of my finite subjectivity flailing around in an infinite reality.

With my memory transformed in this way, it has been possible to harmonize my current understanding of the experience with other elements of my life. Its revelatory nature remains intact. But there is no need to repeat the experience; taking the drugs prescribed to me no longer feels like a regrettable trade-off. The experience stands as seemingly indefeasible evidence for the following background condition to my existence: namely, the sense that reality—both insofar as it seems to be my own reality and the reality of things distinct to me—both outstrips any attempt at comprehension of which I am aware and appears to be such that no attempt by a finite being such as me, or community of such beings could ever do that. However, this needs to be qualified in a crucial way. For it is not a dogmatic commitment. I can't see how my existence could be rendered intelligible conceptually, but I don't take that to be grounds for taking this to be the final word.

This is then combined with a way of managing living in a reality which is inhabited by what appear to be people who do not live with this background condition. Here I adopt the customs of those people to the extent that is needed in order to get by. There is a lot that could be said about what "to get by" means to me. It at least requires that I allow the background sense of things being unintelligible to remain apparent to me as the most truthful-seeming sense of reality that I have. However, it is something that I have also rendered existentially consistent with engaging in other practices that might seem to be at odds with it.

Thus, I am happy to employ the kind of thinking that has the possibility of representing things as they are independently of their representation as its goal, and which takes it to be the case that there are better and worse ways of approaching that task. And part of this includes taking seriously the differences between claims to knowledge and claims which do not warrant this status, as well as being interested in the difference between beliefs which are more or less probable in cases where claims to knowledge appear innapropriate. Furthermore, I am happy to take seriously the idea that my behavior should be subject to the claims of morality. And in both cases, I am happy to rely on a distinction between true and false claims.

The position is one which does justice to what one might think of as a sceptical disposition, but it does not involve a commitment to the impossibility of knowledge. It should also be noted at this point that there are some with whom it seems more important to me to get by with than others; and aligning myself with the epistemic and moral norms that seem to govern the lives of those people is a crucial part of that. It is a messy business; the messy business of my day to day attempt to live what seems to be a good life. It is a domain in which I take myself to have no particular expertise and in which I try to be open to all the help I can cope with receiving. But to reiterate, always in the background, is the only thing that has ever been able to help me get by with existence as a seeming whole given my manic past—namely, that there is just too much reality for any of these customs to be revealing things as they are in themselves; and that none of the customs, even the custom of Pyrrhonian scepticism that I have appealed to in order to get by in the wake of mania, can be taken to be the final word.

Sofia's story

I can't say for certain how long I've been Mad, but it goes back to my childhood. Unlike Paul, I've never received a precise diagnosis, but I sometimes say that I have "schizo-something"; I don't remember precisely when, but in the late nineteen nineties, a psychiatrist said that although I don't tick enough boxes for schizophrenia, I'm likely somewhere on the spectrum. I have also described myself as "having some kind of psychosis thing", but mostly, I make do with "Mad".

When first hearing Paul talk about his experiences and how Pyrrhonian scepticism had helped him, I thought this could certainly not help *me.* I objected that whereas Paul's Madness seemed awesome and therefore understandably tempting, my own was nothing but horrible.

In my papers and presentations (e.g., Jeppsson 2021, 2023a, 2023b, 2023c), I write and talk of *The Mainstream World* and *The Demon World* respectively. The former is the world most people inhabit and share; the latter, as the name implies, a hellish nightmare world filled with murderous demons. Whereas most people trust *The Mainstream World* implicitly—it's so obvious to them that they don't even have a special name for it, it's just *reality*—it always seemed unnervingly flimsy to me. Sometimes the cracks would be showing, through which my supernatural enemies might slip through. Sometimes *The Mainstream World* would flutter and fall apart altogether, and I would be plunged down to what lies beneath.

Throughout my life as a Madperson, I've tried different strategies for dealing with this horror show. I think they can be roughly divided into three groups: Medication, Jamesian strategy, and Pyrrhonism/sceptical strategy. Medication is the obvious one, and the primary help you're offered by psychiatry and the mental health system. My problem: bizarre and terrifying illusions and hallucinations. The solution: give me psychotropic drugs that make them go away. For many years, I was on the antipsychotic drug Haldol, the sleeping pill Propavan, and occasionally the benzodiazepine Xanax. However, dealing with my psychosis by taking antipsychotics was more complicated than most people realize.

Sane people take *The Mainstream World* for granted. They might even find it *impossible* to truly doubt it. They might—like David Hume—entertain sceptical arguments, perhaps feel briefly shaken by them from time to time, but they soon return to trusting that the world is the way they always thought it to be. Of course, sane people sometimes change, e.g., their ideological or religious views in a way that may feel dramatic enough to the person concerned, but throughout these changes they never experience or believe in anything like my several layers of reality or hostile demons coming up from the world beneath. When you take *The Mainstream World* for granted like this, antipsychotics might seem like an obvious and simple solution for the kind of frightening experiences that I've dealt with.

It wasn't so simple for me. When I first became a psychiatric patient, I was genuinely uncertain of whether antipsychotics would suppress frightening illusions and hallucinations, or blind and deafen me to a *Demon World* and demons that were really there, making me much more vulnerable. Decades later, as a philosopher, I can explain how all scientific arguments for what is and what isn't possible presuppose *The Mainstream World,* and that philosophical arguments attempting to show that people are justified in trusting it, in turn, presuppose that the trust is already there. Unfortunately, there's no scientific proof or philosophical argument relevant to the poor Madperson who already finds themself floating between realities, doubting and questioning everything. Back then, I had yet to study philosophy, and couldn't put all of this into words. Nevertheless, I noted that whenever a doctor tried to explain to me that my demons were unreal and the meds would help me, there was something circular or question-begging about their arguments, which left me feeling frustrated and profoundly unconvinced.

Thus, I had to supplement the medical solution by what I've later come to call the Jamesian strategy, after philosopher and psychologist William James. James (1896/2010) argued that there are rare circumstances in which we lack sufficient evidence one way or the other, and yet a neutral suspension of judgment isn't an option; the stakes are *high* and we must believe *something.* In situations like these, he said, it makes sense to *choose* what to believe. I made a pure leap of faith and *choose* to believe that *The Mainstream World* is the sole reality, *The Demon World* and its inhabitants are just figments of my psychosis, and taking the pills therefore made sense.

However, choosing what to believe is hard; you can't sustain a belief by pure willpower for long. Fortunately, I didn't have to. My then-psychiatrist tried a few different medications before striking gold with Haldol, which for a long time worked very well. Once the pills made *The Mainstream World* stabilize around me, it eventually came to seem obvious to me, perhaps as obvious as it seems to sane people.

From time to time, I would think myself cured for good, quit my meds, and sail on for a while, until some triggering experience (for instance, changing environments and going abroad, or something more traumatic) sent me flying back to *The Demon World* again. I once again had to *choose* to believe in *The Mainstream World*, psychiatry and its pills, and get back on them until the world restabilized and my trust in it returned.

However, Haldol eventually began losing its desired effect on me, while simultaneously giving me increasingly nasty side effects. I had to rely more and more on Xanax not to completely freak out, despite knowing full well what a dangerous drug it is. Eventually, I was also given the beta-blocker Propanolol, but this last one never had any effect on me, not even at dosages of 100 mg at a time. *The Mainstream World* was flimsy again, the demons pushed through more and more often, and my implicit trust in the *Mainstream* eroded.

In hindsight, I realize that I actually felt betrayed by psychiatry as a whole, and even betrayed by the Haldol pills themselves. Thi Nguyen (2022) has written on the similarities between trusting another person and trusting an object or machine, and reading his paper gave me an eureka moment—*that's* why I had such intense emotions about my medication: I used to *trust* it, but then it *betrayed* me! As a psychiatric patient, you're taught to trust your medication, and to believe that if only you hold up your end of the bargain by conscientiously taking the pills as prescribed, the pills will do their job and keep you sane. But Haldol, eventually, didn't.

After I became friends with Paul—introduced to me by a colleague as a "fellow Mad philosopher, you might have much to talk about"—he told me how he had found Pyrrhonian scepticism helpful for dealing with his own non-standard experiences, but at the time, I wasn't ready to listen. I still yearned for the days gone by when *The Mainstream World* felt stable and firm enough to be trusted, and when I experienced something like sanity. I tried to go back to this state by a continuous, Jamesian effort of will alone, but it was hopeless.

Eventually, I went to see a psychodynamically trained therapist for my own money. This was in 2019, at which point the public health care system had long relied on drugs and cognitive behavioural therapy. But I was certain I needed to talk things out with someone who was willing to go deeper, who would be open-minded and willing to explore where the discussions would lead. Finally, I found a therapist who did something other than CBT and job training. After a few sessions, she said that I seemed too hung up on what's *Mad* and therefore *bad* and *must not be done* instead of simply utilizing whatever strategies and coping mechanisms that help me feel better and prevent me from freaking out. This was a real eye-opener for me; I hadn't realized, before, how much internalized stigma I was carrying around. I do think there's a connection here between "strategy" in the sense of a consciously chosen way to *deal* with your Madness, and Madness as itself being a strategy in Garson's sense. If you embrace the latter—if you see it as fulfilling some *purpose* and being helpful in *some* ways even as it may harm you in others—you might be more open to the idea that a consciously chosen strategy need not be bad just because it seems to be, in itself, quite Mad.

Usually, when people talk about "stigma against mental health conditions" and how we should fight said stigma, it's construed as people being ashamed of saying that they have a mental health condition in the first place, and/or people being shamed for taking meds. And sure, those are aspects of stigma. But another aspect is the pressure people like me feel to construe their Madness as running less deep than it does. I used to insist that I obviously know what's real or not, I just suffer from a little brain dysfunction, that's all. It's like asthma or diabetes except in the brain, nothing to see here, move along. Now, encouraged by my therapist, I finally admitted to myself that I often *don't* know what's real—but that's okay, as long as I still manage to roll with things and live my life.

And so, we arrive at my third and most fruitful strategy for dealing with Madness: the one I label Pyrrhonian, after the philosophical school of Pyrrhonian scepticism.

Now, this name might not be entirely apt after all. The ancient philosopher and physician Sextus Empiricus (1976) didn't write about shifting between two different worlds, and then remain neutral about whether one or both were real. Rather, he writes about accepting that there are always different perspectives from which to see things, and there are always counter-arguments against as well as pro-arguments for our beliefs. Nevertheless, he stresses that when we come to accept this, we can reach a peace of mind not possible for the person frantically trying to determine what's true or not. And he does bring up Madpeople in his writings: Sextus writes that even if it were true that Madpeople had a different balance of humors than sane people do, we wouldn’t have any independent proof of which humor balance makes you see the world as it really is and which distorts it.

The idea of finding peace by accepting that I can’t know what’s real or not resonated with me. I further realized that by now, I have reason not to fear the demons either way. Either the demons can’t kill me because they’re not real, or they very likely won’t kill me because they’ve been stalking and threatening me for decades and I’m still alive, so those threats seem pretty empty. This either-or thought comforts me in a way that insisting on just the first part—they’re not real! Not real!—can’t do. I also realized, with the help of my therapist, that I don’t need to determine what’s real or not to know *what to do.* I have ways of dealing with my demons—talking to them, engaging in little protective rituals—that are justified if they’re real, and also justified if they’re not; it keeps me from spiralling into ever higher stress- and fear levels, and thereby keeps me from a full psychotic breakdown.

This Pyrrhonian strategy has been immensely helpful to me once my meds no longer worked. Nevertheless, it took me even longer to acknowledge that the terrifying *Demon World* held a certain *allure*. I used to think that whereas Paul felt understandably tempted by what seemed like the prospect of vast cosmic insight, murderous demons are wholly scary and bad. I told him that unlike him, I *wanted* to believe, whole-heartedly and without hesitation, that my demons were nothing but illness symptoms and *The Demon World* an illusion. The only reason I eventually came to embrace Pyrrhonian scepticism was because I was out of other options; neither medication nor wilful Jamesian believing worked anymore. But for me, being in a state of florid psychosis felt like being the main character in a horror movie. Who on earth would be *tempted* by that?

Eventually, I came to realize that even if you’d rather be the main character in a nicer kind of movie, simply *being the main character* has a certain allure compared to being one of eight billion bit-players in the regular world. Moreover, regardless of how terrified I’ve been when actively psychotic, I’ve never been *bored.* And finally, even a hellish *Demon World* might seem more manageable on occasion than *The Mainstream World* that most people inhabit.

I’m not sure how common my experiences are, but they’re not unique. I recently met Kay A. Subijana at a conference, who’s also had terrifying psychotic experiences, and agreed with me about the last point. You suddenly find yourself the main character of a story which is incredibly scary but *simple*. In my case, pursued by demons who try to kill me (why? They’re weirdly attached to me and evil—my subconscious never built up more backstory or personality for them than that), and I must avoid being killed. That’s it. Kay hasn’t, of course, been through the exact same experiences as I have, but they’ve found themself in the midst of similarly scary but *simple* narratives when psychotic.

The problems I face in *The Mainstream World* may rarely concern the prospect of my immediate murder, but they can have a sprawling complexity which is terrifying in itself. Moreover, whereas I have two standard options when I want to avoid murder by demons—try to protect myself or flee—the problems of the Mainstream World are often such that it’s hard to see what the best strategy would be; there might not even *exist* any solutions. Even terrifying kinds of Madness can serve the escape function that Wigan and Podvoll talked about.

I have found that admitting to feeling tempted in the first place makes temptation easier to handle. When I regarded any pull felt as an incomprehensible mental illness symptom, there was nothing I could do about it except medication or resistance through brute willpower. Once I admit that there are reasons to feel tempted, I can rehearse my reasons for and against. If I slide into florid psychosis, I’ll feel at the centre of the world, it will be terrifying but *exciting,* and I’ll have less complex problems to deal with. However, I have important responsibilities to and relationships with people I care deeply about, and I need to stay connected to *The Mainstream World* to preserve them. Moreover, regardless of what kind of *Mainstream* mess I find myself in, it will likely have grown bigger and messier by the time I return if I first take an extended psychosis break. The temptation to go Full Mad can still be hard to handle, sometimes—in particular since my Madness has grown less terrifying and more benign in later years. Nevertheless, it’s more doable once I’ve admitted to myself that Madness can serve a purpose and offer an escape from *Mainstream World* problems.

Conclusion

Our aim with this paper has been to sketch the way in which we have individually embraced a Pyrrhonian strategy for dealing with our respective Madnesses and their allure. However, in closing we would again like to make it clear that we remain pluralists about the place of self-views, world-views and narratives in coping with Mad existence (how could it be otherwise, given that we both find value in Pyrrhonian scepticism?) Indeed, as should be evident from our discussion, there are differences between the ways that we ourselves understand the significance of Pyrrhonian scepticism and employ it in our own lives. As a result, we welcome the contribution that Justin Garson's rich analyses make available to those, Mad or otherwise, who are struggling to think about Madness. If other Madpeople find it helpful to adopt a pure dysfunction view, we certainly do not wish to argue that they are wrong. Nevertheless, given our own journeys and the place the Pyrrhonian scepticism has come to play in those, we think it is crucial that this approach doesn't become too dominant. The different perspectives that Garson details in his book—of Madness as a strategy, an escape, a temptation—offers important complements. Many Madpeople, their friends and families, as well as clinicians, would do well to at least contemplate alternative perspectives from time to time; and perhaps for some the sceptical perspective will seem like an appealing option.

Acknowledgments

We would like to thank the audience at the conference Too mad to be true II and following people for their helpful discussion of the issues and comments on earlier versions of the paper: Katherine O'Donnell, David Mordecai, Rita Ritunnano, Amia Srinivasan, Henry Straughan, and David Velleman.

REFERENCES

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. 5th ed. Washington, D.C: American Psychiatric Association.

Garson, Justin. 2022. Madness. *A Philosophical Exploration*. New York: Oxford University Press.

James, William. 1896/2010. *The Will to Believe, and other Essays in Popular Philosophy*. Auckland, New Zealand: The Floating Press.

Jeppsson, Sofia. 2021. "Psychosis and Intelligibility." *Philosophy, Psychiatry and Psychology*. 28 (3): 233-249.

———. 2023a. "My Strategies for Dealing with Radical Psychotic Doubt: A Schizo-Something Philosopher's Tale." *Schizophrenia Bulletin* 49 (5): 1097–98. https://doi.org/10.1093/schbul/sbac074.

———. 2023b. "Radical Psychotic Doubt and Epistemology." *Philosophical Psychology* 36 (8): 1482–1506. https://doi.org/10.1080/09515089.2022.2147815.

———. 2023c. "Exemption, Self-exemption, and Compassionate self-excuse." *International Mad Studies Journal*. 1 (1): 1-21

Nguyen, Thi. 2022. Trust as an unquestioning attitude. In *Oxford Studies in Epistemology* vol. 7, edited by T. S. Gendler, 214-244. Oxford: Oxford University Press.

Podvoll, Edward M. 1991. *The Seduction of Madness*. London: Century Press.

Sextus Empiricus. 1976. Vol. 1. *Outlines of Pyrrhonism*. Translation by R. G. Bury. Cambridge, M.A.: Harvard University Press.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 2 | 133-153
https://doi.org/10.31820/ejap.21.2.3

INTO THE DEEP END: FROM *MADNESS-AS-STRATEGY* TO *MADNESS-AS-RIGHT*

Miguel Núñez de Prado-Gordillo[1]

[1] University of Rijeka, Croatia

Discussion – Received: 29/08/2024 Accepted: 05/11/2024

This paper is part of a book symposium on Justin Garson's *Madness: A Philosophical Exploration* curated and edited by Elisabetta Lalumera (University of Bologna) and Marko Jurjako (University of Rijeka)

ABSTRACT

A central notion in Mad Pride activism is that "madness is a natural reaction" (Curtis et al. 2000, 22). In Madness: A Philosophical Exploration (2022), Justin Garson provides a compelling exploration and defence of this idea through the book's central concept: madness-as-strategy, i.e., the view of madness as "a well-oiled machine, one in which all of the components work exactly as they ought" (1). This contrasts with the dominant view in 20th- and 21st-century psychiatry, madness-as-dysfunction, which understands madness as a failure of function. The paper provides a critical analysis of the notion of madness-as-strategy as a political tool, pointing out its main virtues and limitations in terms of Garson's overarching political project: to carve out the conceptual landscape of madness in ways that pay tribute to mad people's own perspectives. The analysis draws on two central commitments of contemporary neurodiversity theory: a) its relational-ecological model of cognitive (dis)ability; and b) its non-essentialist, sociopolitical critique of the "normalcy paradigm". I argue that these two insights contribute to both expand the applicability of madness-as-strategy and highlight its limitations as a tool for the political struggles of mad, cognitively divergent, and mentally ill or disabled people. The paper concludes by outlining a way to move beyond both madness-as-dysfunction and madness-as-strategy, toward what I call madness-as-right.

Keywords: philosophy of psychiatry; conceptual explication; mad studies; neurodiversity paradigm; madness-as-dysfunction.

Correspondence: m.nunezdeprado.gor@ffri.uniri.hr

Introduction

In his 2017 song "YAH.", Compton-born and raised rapper Kendrick Lamar claims to be "diagnosed with real [n-word] conditions". The theme behind this verse is a common one in Lamar's production (e.g., his 2012 album *Good Kid, M.A.A.D. City*): that his struggles with mental health are the result of a natural, adaptive response to a mad environment. As Mad Pride founder Pete Shaughnessy puts it:

> I see life as one big swimming pool. Some of us are thrust into the deep end and we manage to survive. We make our way down to the shallow end, where it's easy, boring. The people there are scared of the deep end, scared of the unknown, so they shun people like me and call me MAD. Madness is a natural reaction. (Shaughnessy 2000, 22)

Justin Garson's (2022) *Madness: A Philosophical Exploration* provides an engaging, thorough, and compelling exploration of this precise topic. Its central concept, *madness-as-strategy*, conveys both Lamar's and Shaughnessy's main insight: that madness is the expression, under certain circumstances, of "the working out of a hidden purpose; instead of a defect, (…) a goal-driven process, a well-oiled machine, one in which all of the components work *exactly as they ought*" (1). This concept stands in contrast to a more common way of understanding madness: what Garson labels the *madness-as-dysfunction* view, which identifies it with a failure or breakdown in some internal machinery. Madness here "represents the failure of the system to achieve its natural end" (1). This is the key contrast that the book focuses on; one that is orthogonal to the more classical debate between biogenic vs. psychogenic approaches to mental health, concerning whether mental health should be conceptualized in somatic (e.g., neural) or mental terms. Rather, it is *teleology* vs. *dysteleology* which interests the author: that is, whether madness can or should be seen as the product of a strategy, a purpose, a well-functioning mechanism—whether mental or somatic—or as its failure.

According to the author, madness-as-dysfunction represents the dominant way of thinking about madness in contemporary mental health science and philosophy. So entrenched the association between madness and dysfunction is, Garson thinks, that some have come to view it "as a matter of logical necessity" (11; e.g., Boorse 1976; Wakefield 1992), rendering any alternative concept of madness "almost unthinkable" (248). The primary goal of the book is in this sense straightforward: to question the often-assumed unquestionability of the madness-as-dysfunction perspective by reasserting the conceptual plausibility of madness-as-strategy, as expression of "a hidden telos" (3).

To do so, Garson uses an engaging mix of historical analysis and conceptual engineering. The book's method consists in analysing different theories of madness throughout history in terms of the proposed distinction between teleology and dysteleology—from the clash between ancient Greece conjurers and the first Hippocratic thinkers to the Christianization of madness as both punishment and salvation characteristic of the Middle Ages; from the progressive secularisation of madness throughout the early modern period, culminating with the Kantian understanding of it as a breakdown of reason, to the discussions between psychogenic and biogenic approaches characteristic of 20th and 21-st century psychiatry. In doing so, it explores the different shapes that the teleology of madness has taken throughout history: from "a divine mandate [to] "a mysterious vital principle in nature; (…) an unconscious idea driving toward fulfilment; (…) the goal-directedness of the organism; [or] a Darwinian adaptation" (13). Despite its historical outlook, however, the book "is not a work of history, but an exercise in concept building" (3). Here it aligns with recent approaches to the philosophy of psychiatry that adopt an explicationist or "engineering" methodology to "craft" new concepts fit for specific purposes, rather than merely analysing existing ones (e.g., Biturajac and Jurjako 2022). Specifically, the book doesn't aim to faithfully reconstruct the dialectics of the different ways of thinking about madness throughout history, but to extract from different theories "a teleological core, an attempt to think of madness as a strategy for accomplishing a goal" (3), to make room for a functional view of madness *today*.

Note, however, that although the author favours madness-as-strategy, the goal is not to defend it for the sake of it, but to *use* it in an attempt to crack open the established, almost self-evident consensus around madness-as-dysfunction. In this sense, the book's ultimate aim is that, by historicizing assumptions about madness, it helps carve out the conceptual space for new ways of thinking about it that transcend *both* madness-as-dysfunction and madness-as-strategy.

My main goal here is to contribute to this effort. After explaining the structure of the book in section 1, section 2 provides a critical analysis on the concept of madness-as-strategy, its merits, limitations, and possibilities for future development. Here I align with the book's methodology, as well as its political ambitions and the new ways of thinking it encourages, which start from taking seriously the perspectives of those at the "deep end" of mental health science: those who identify as mad, as survivors, as cognitively divergent; but also, those who identify as (ex-)patients, as mentally ill, or cognitively disabled. Specifically, I claim that the emerging neurodiversity paradigm (Chapman 2023b; Walker 2021) offers key conceptual tools to integrate these different—and sometimes conflicting (Spandler, Anderson, and Sapey 2015)—modes of conceptualizing madness and mental health from the "deep-end" perspective; and it does so in a way that invites us to move beyond *both* madness-as-dysfunction and madness-as-strategy, toward what I will refer to as *madness-as-right*. In this sense, the paper seeks to contribute to ongoing efforts to develop a collective pool of conceptual resources integrating insights from neurodiversity theory, mad studies, and disability studies (Graby 2015; McWade, Milton, and Beresford 2015); efforts that, like Kendrick Lamar's recent *The Pop Out* concert—where the artist transformed his historic beef with Drake into a momentous display of unison for the Black community in Los Angeles—seek to foster unity and comradeship while still acknowledging diversity.

1. The hidden telos of madness

The book is structured in three parts, which roughly divide pre-modern, modern, and contemporary views of madness. The first part of the book, "The Dual Teleology of Madness", covers pre-18th century views that share a common underlying assumption: that madness, whether construed as failure or strategy, takes place within a larger divine teleology. It is always a result of divine intent, an instrument of divine justice–as well as redemption in Christianity, hence its *dual*teleological nature–executed by direct divine intervention or preconfigured in how God designed the world in the first place.

This is why Garson reconstructs the main oppositions throughout this period as not primarily between supernatural vs. naturalistic explanations—as contemporary medicine textbooks often portray it—but between teleology and dysteleology; between the madness-as-strategy view of pre-Hippocrates healers, who characterized it as a divine punishment, and the madness-as-dysfunction tradition installed by Hippocratic physicians, where it results from inner humoral imbalances (Chapter 1); between the understanding of madness as demonic possession, characteristic of Christian exorcists and witch-hunters during medieval ages, and the attempt to reintroduce the Hippocratic framework by witch-sceptic, Renaissance-minded physicians like Jorden (Chapter 2). But even Hippocratic, dysteleological views of madness are, during this period, only carved out against a broader divine teleological framework. To be sure, Garson observes a progressive naturalisation of divine teleology, and therefore of madness, throughout this period. This is already visible in 17-

th and 18th-century physicians like Burton (Chapter 3) or Cheyne (Chapter 4), who viewed madness as a natural, inevitable consequence of how God designed the causal order in the first place. Here, madness is a condition that *naturally* follows from our "freely chosen and wilful misuse of our God-given faculties" (49), as in Burton's case; or a disorder of the nerves resulting from sustained habits of intemperance, an offence to God's providence, in terms of Cheyne.

The second part of the book, "Madness and the Sound Mind", mainly tackles the rise of madness-as-dysfunction during 18th and 19th centuries. The key characteristic of this period is the association between madness and the sound mind: to fully understand madness, it is crucial to first understand what universal mental faculties characterize well-functioning mentality; madness just is their *breakdown*.

Garson sees in Kant a most articulate early expression of this view; for every faculty of the sound mind, a variety of madness that results from its dysfunction (Chapter 5). Madness no longer reveals any hidden, divine telos. Still, there are remnants of teleology in madness. For Kant, following Locke, because even the gravest forms of insanity still exhibit a "*systematizing* tendency" (89), i.e., they organize around a somewhat coherent—even if fundamentally distorted—inferential whole. For Haslam (Chapter 6), apothecary to Bedlam, the so-called first psychiatric institution, because this participation of madness in reason reveals its ultimate purpose: "to dissimulate reason in order to perpetuate its own existence as madness" (95). This dissimulation function is also highlighted by Wigan, who thought of madness as resulting from the inherent duality of our mind-brains (Chapter 7). Dissimulation here, however, is not a means for deceiving others, but *oneself*: a way in which the "sick hemisphere" might gain ascendance over the healthy one—for instance, to *cope* with an otherwise unbearable reality.

Heinroth (Chapter 8), although assuming a thoroughly Kantian, madness-as-dysfunction framework, also makes room for at least some purposive form of insanity-as-coping, "as a way of retreating or withdrawing from a lifetime of suffering, tragedy, ridicule, and disdain, and entering into a kind of dream world" (130); a view of madness that is also the conceptual cornerstone of 20th-century psychoanalytic views of schizophrenia, such as those of Fromm-Reichmann or Sullivan. This coping perspective is tightly connected with a view of madness as some sort of "healing journey", and hence of therapy as a form of shepherding the person along this journey. This is the framework in which Pinel's moral therapy must be understood (Chapter 9). Pinel, according to Garson, pushes beyond the Kantian framework: madness is paradigmatically purposive: just like fever is the body's own natural mechanism for healing, madness "is a healing and salutary movement of the mind" (154). Hence the role of the moral therapist: not to interfere with madness, but to facilitate it, "to allow it to reach its natural end" (155).

The second part however finishes with German imperial psychiatry (Chapter 10), which reinstalls madness-as-dysfunction in all its force by "fusing together (…) two doctrines—that madness is biological, and that madness is dysfunctional, or more concisely, that mental disorders are biological dysfunction" (160). Griesinger, the so-called father of biological psychiatry (Shorter 1996), emphasizes the former in his "biologization" of Kant: madness is a dysfunction of brain processes. Yet Griesinger still sees traces of teleology in madness; for instance, in his characterisation of delusions as "wish fulfilments". By contrast, Kraepelin's naturalisation of psychiatry emphasizes its definitive purge from teleology: madness is *necessarily* dysteleological—and if we can spot any trace of teleology in it, then it is not true madness, but mere *malingering*.

Finally, the third part, "Madness and the Goal of Evolution", covers 20th- and 21st-century perspectives; a period marked by an oscillating, yet largely unnoticed tension between teleology and dysteleology.

At least until the 1960s, madness-as-strategy is somewhat predominant due to the influence of psychoanalytic theory. Freud opens the century restoring teleology firmly at the core of psychiatry: madness is *always* functional (Chapter 11). Specifically, Freud's madness is a dual strategy for the control of forbidden, self-destabilizing desires: it keeps them unconscious, safeguarding one's self-concept, while at the same time offering a (deviant) way of fulfilling them. This leads to an "anti-Kantian", "anti-Kraepelinian" classification scheme, implemented in the first edition of the *Diagnostic and Statistical Manual of Mental Disorders* (APA 1952): one which classifies varieties of madness (e.g., psychotic, neurotic, and personality disorders) in terms of the "different strategies that the mind uses to fulfill its twofold function of keeping forbidden desires out of consciousness while orchestrating their deviant fulfillment" (188). Goldstein, according to the author, *biologizes* Freud by placing the analysis of disease and disorder within a *holistic* philosophy of biology, which takes the essential *self-actualizing* goal of whole organisms as its starting point (Chapter 12). Working mainly with brain-injured veterans, he conceptualizes their symptoms as primarily a self-stabilisation strategy, deployed via restructuring the environment in ways that compensate for the anxiety-inducing, de-stabilizing experiential consequences of the injury.

This relational-teleological characterisation of madness as a mode of engagement with the world, as well as the emphasis on its creative, world-changing power, is also present in Laing's redefinition of madness as a revolutionary tool (Chapter 13). Following the insanity-as-coping intellectual tradition initiated by Heinroth and later continued by Fromm-Reichmann and Bateson's double-bind theory, Laing views madness as an adaptive response to disturbing double-binding patterns of communication within the family structure. For Laing, however, the origin of such disordered patterns must be traced back to the larger political order. Thus, unlike their intellectual forebears, Laing and other so-called "anti-psychiatrists" from the 1960s counterculture see madness as no mere retreat from the world: it is a revolutionary negation of it, an "assertive refusal to participate" from the (in)sane, normal, capitalist social order (213). At least *good, true,* in fact, *sane* insanity—which counterculture thinkers like Deleuze and Guattari contrast with the "false", "useless" madness of the "gibbering lunatic" (215)—has this revolutionary function; like Pinel, the job of the psychiatrist is to shepherd the mad person; not "back to normal" anymore, though, but toward realizing their revolutionary potential.

However, madness-as-strategy would progressively recede during the 1970s, with the advent of the second-wave biological psychiatry and its first "neurotransmitter imbalance hypotheses". This brought the dissolution of any differentiation between "good" and "bad" madness: madness, in all its varieties, would increasingly be considered the result of inner dysfunction. This neo-Kantian, neo-Kraepelinian *deteleologization* of madness already begins with the *DSM-II* (APA 1967) and finds its maximal expression in the *DSM-III* (APA 1980) and the various attempts to cast a workable notion of *dysfunction* (Chapter 14). For this is its central concept: against the common misreading that the DSM-III established a *biogenic* or *bio*medical regime, the author reminds us of the "atheorical", cosmopolitan spirit that guided its development. Its core feature rather is its answer to the *boundary problem*, i.e., its definition of madness as (inner) dysfunction—whether biological or psychological—to distinguish it from mere social deviance. However, its own notion of dysfunction, influenced by Spitzer and Endicott's operational proposal, is just too vague, leaving the relevant domains of functioning open to culture-specific understandings. To secure its universality, psychiatrists turn to evolutionary theory. Problems with Kendell's initial definition as any condition which intrinsically places the individual at "biological disadvantage" led to Klein's definition of disorder as "deviation from evolved design" (247), i.e., from what evolutionary contingencies selected the human mind and body parts to do; a definition that Wakefield's harmful dysfunction analysis would later convert into a conceptual necessity, and which the recent RDoC framework, in the latest and most "systematic and unforgiving" (230) application of Kantian dysteleological nosology, takes as its fundamental axiom for classifying mental disorders.

However, the book closes by revealing an internal tension within this "Darwinization of madness" (Chapter 15); one which at the very least risks undermining madness-as-dysfunction. As contemporary adaptationist hypotheses of psychiatric conditions show (e.g., Nesse and Williams 1994), evolutionary theory provides us with good reason to see in madness a product, not a failure, of evolved design; to see psychiatric conditions like depression, anxiety, or delusions as the result of *mismatches*—that is, evolutionary adaptations that are no longer beneficial in current environments—or even adaptations that are still serving their original functions. This, according to Garson, "forces a teleological reorientation of the entire discipline" (252): one that highlights the historicity of madness-as-dysfunction, as well as its actual tension, rather than kinship, with evolutionary theory, not to reject it, but to question its status as "a silent default in approaching the mad (…), "to identify and expose [it] as merely one style of thinking, and to force it to coexist with other styles of thinking" (260-261).

2. Beyond madness-as-strategy: Madness-as-right

Garson's *Madness* has multiple virtues, some of which are set out right in the introduction. Firstly, I think the book's proposed reorientation of the history of psychiatry, i.e., its focus on teleology vs. dysteleology, rather than the more usual contrasts between "somatic" and "mental", or "biological" and "psychological", is extremely illuminating. I think this new axis of analysis is not only original and refreshing, but crucial to fully understand the conceptual structure and historical roots of the so-called *medical* model of mental distress—so often wrongly conflated with the *bio*medical one. As the book very clearly shows, it helps to dispel common misunderstandings of the main conceptual transformations reflected in and partly brought about by the DSM-III—even the DSM-II, as Garson convincingly argues—and the subsequent evolution of contemporary psychiatry. This surely leaves some questions unanswered; for instance, how central should we take dysteleology to be for medicalization? Taking dysteleology as the ultimate *hallmark* of what makes a model "medical" would seemingly—and I think wrongly—suggest that psychoanalysis was not, after all, a genuinely *medical* approach; a claim that would be, at the very least, difficult to reconcile with most 20th-century psychoanalysts' self-perceived status. Nonetheless, Garson's proposed redirection helps us to uncover the deep conceptual affinity between seemingly, but only superficially opposite approaches to mental health science, e.g., between DSM's syndrome-based and RDoC's bottom-up, dimensional approach to nosology; between classic biomedical "chemical-imbalance", "magic-bullet" approaches to psychopharmacology and the new, self-avowedly revolutionary psychedelic psychiatry; or between biological psychiatry and competing psychogenic disciplines, such as clinical psychology (at least in its traditional cognitivist versions). In this sense, the teleology-dysteleology distinction is a crucial addition to our conceptual toolkit.

However, the book's most important contribution is, as the author himself notes, primarily political. Madness-as-strategy, its central concept, is not only a theoretically sound analytical tool, but also a political instrument that Garson systematically uses to point out the historicity of madness-as-dysfunction; not just for the sake of historical and conceptual accuracy, but to help reshape the conceptual space of madness in ways that pay tribute to mad perspectives themselves. The book in this sense contributes to recent efforts at providing conceptual support for the long-standing struggle of mad, survivor, neurodivergent, and related collectives to put their own expertise and perspectives in value, to reclaim their space in mental health science and politics (Adler-Bolton and Vierkant 2022; Chapman 2023b; Frazer-Carroll 2023; Rashed 2019; Walker 2021). It is therefore a contribution to the political struggle of those traditionally relegated to the "receiving end" of psy-services; or those "thrust into the deep end", as Pete Shaughnessy would put it.

Here I want to delve deeper into this issue, to push forward in this same direction. In that sense, this paper takes the book's political ambitions at face value. As Garson himself points out, however, I think that moving forward in this direction requires transcending not only madness-as-dysfunction, but also madness-as-strategy. The main reason why I think so is that the deep end of mental health science and politics is primarily characterized by a rich—and sometimes conflicting—multiplicity and diversity of first-person perspectives, which neither madness-as-dysfunction nor madness-as-strategy can properly accommodate (Spandler, Anderson, and Sapey 2015). My starting point is contemporary neurodiversity theory, which I think offers various key insights to develop a conceptual framework that connects and reconciles intersecting critical views of mental health.

Very briefly, neurodiversity theory is an emergent field of study that aims to integrate and develop the theoretical architecture of the neurodiversity movement. Born in the 1990s from collective discussions within the autistic community (Botha et al. 2024; Rosqvist, Chown, and Stenning 2020), the movement has been increasingly applied to the analysis of other developmental conditions, such as ADHD or intellectual disabilities; furthermore, the concept of *neurodivergence* is increasingly applied to a broader range of conditions that involve some departure from prevailing standards of "cognitive normality"—including bipolarity, obsessive-compulsivity, depression, schizophrenia, borderline and antisocial personality, etc.[1] (see Chapman, 2019, 2023b; Hoffman, 2019; Jeppsson, 2023; Rosqvist, Chown, and Stenning 2020; Walker, 2021).

The movement is theoretically articulated around the emerging *neurodiversity paradigm* (Walker 2021), whose core commitment is the critique of the *default pathologizing*, as well as the *default normalizing* of divergent cognitive styles (see also Chapman, 2023b). Unlike traditional psychiatric models that equate deviation from "neuronormative" standards with inner dysfunction, the neurodiversity paradigm sees cognitive diversity as a natural and valuable part of human variation, along with other forms of biodiversity. Differences in sensorimotor and cognitive functioning are not necessarily "deficits"; in fact, they might bring both individual and collective *advantages* in certain contexts over more neuronormative modes of functioning (Chapman 2021; Crompton et al. 2020; Dwyer 2022; Sedgwick, Merwood, and Asherson 2019). At the same time, the paradigm also opposes normalizing, "anti-disability" discourses, found for instance in other traditional critical perspectives that take an abolitionist perspective on psychiatric categories (e.g., Szasz 1961), which often question the existence of genuine cognitive differences or downplay their disabling nature (Carel 2023; Chapman 2023a; Milton 2014; Walker, 2021).

For the purposes of this paper, the neurodiversity paradigm's most significant contributions lie in a) its relational-ecological understanding of cognitive (dis)ability; and b) its sociopolitical, non-essentialist understanding of mental categories. Firstly, neurodiversity theorists reject "inner deficit" or "inner dysfunction" views of cognitive divergence (Chapman 2021; Milton 2012; Walker 2021). In line with the social model of disability, the neurodiversity paradigm construes the difficulties faced by cognitively divergent people as the result of a *mismatch* between their cognitive traits and the socio-material environments they navigate. Cognitive (dis)ability is here understood in a fundamentally *relational* way: it is not the result of

[1] A related development concerns the inclusion of mental disorders within the scope of the neurodiversity movement; from this perspective, the concept of *mental disorder* would not be anti-thetical to that of *neurodivergent*, but a subspecies of cognitive divergence—along with other non-pathological forms of neurodivergence (Chapman, 2023b; Hoffman, 2019; Walker, 2021). Although still a matter of debate within the movement (see Kapp, 2020), this paper aligns with this expansion of the neurodiversity framework. I will thus use the term "neurodivergent" as originally intended by its creator, Kassiane Asasumasu, who coined it with the explicit inclusive aim of encompassing "any significant divergence from dominant cultural norms of neurocognitive functioning" (Walker, 2021, p. 47), including mental disorder.

inner deficits, but the result of a failure to design our worlds in ways that accommodate different sensorimotor, cognitive, and behavioural dispositions. Furthermore, the neurodiversity paradigm advances an *ecological* model that takes into account not only individual, but also *collective* cognitive functioning and flourishing (Chapman 2021; Jurgens 2023; see also Hoffman 2017). This ecological view articulates one of the founding ideas of early pro-neurodiversity communities: that, just like biodiversity is crucial for a healthy environment, cognitive diversity within human groups might be an adaptive feature for maximizing *collective* thriving and fitness. Failure to accommodate and cultivate this diversity may not only impact cognitively divergent individuals' health and functioning, but also groups' ability to cope with ever-changing environmental demands.

This relational-ecological model has fruitful implications for the concept of madness-as-strategy. Specifically, I think it helps to widen the scope of the concept; it allows us to cast the net wider on the phenomenon of madness from a teleological perspective, at least in two ways. Firstly, it seems to nicely capture mad advocates' emphasis on the need to understand madness, disorder, or disability as a person-world relation, that is, to analyse how a person's material and social environments enable, enhance, or diminish their cognitive functioning and possibilities for flourishing; an insight that Garson emphasizes repeatedly throughout the book, especially in his consideration of Goldstein's holistic understanding of madness as a mode of engagement with the world (Chapter 12) and Laing's analysis of the constitutive impact of sociopolitical dynamics in madness (Chapter 13; see also Cooper 2017). Indeed, the latter points to a notion that Garson only briefly touches upon in the book, and which is central to the ecological model of cognitive functioning developed by neurodiversity scholars: that of collective (dys)functioning. This is the second way in which neurodiversity theory can help expand the scope of madness-as-strategy, by adding a new level of analysis at which madness may exhibit its hidden telos: not only may madness serve a purpose for the individual, but also for their larger social niche. Furthermore, *even if we accepted madness-as-dysfunction at the individual level*—or, at least, madness-as-disability—we may still look at its adaptive role at the level of collective functioning. In line with the 1960s counterculture revindication of madness as a revolutionary tool, conditions like psychosis, depression, anxiety, ADHD, or autism might be reconsidered in light of their potential contributions toward more adaptive, healthier, *saner* ways of social organization; even if, in our current world, this comes with often extraordinary costs for individuals themselves.

But the neurodiversity paradigm also has a deeper, more crucial, yet perhaps not so positive implication for the concept of madness-as-strategy; namely, its questioning of the extent to which we may speak of a natural, universal, or somehow fixed standard of "normal" cognitive functioning in the first place, as well as its usual immediate, almost *a priori* association with notions of cognitive health and flourishing. A running thread throughout Garson's book is that madness-as-dysfunction and madness-as-strategy fundamentally oppose each other on whether madness is a breakdown in cognitive function or rather an expression of "a well-oiled machine, one in which all of the components work *exactly as they ought*" (1). But this points to a hidden premise that both madness-as-dysfunction and madness-as-strategy seemingly share: that there is something like a "well-oiled machine" in the first place with which madness can be compared, some essential assortment of mental functions and capacities that conform a natural or universal standard of *normal* cognitive functioning; a fixed mould into which madness must fit if we are to see purpose, value, and an enactment of human cognitive potential in it. Madness-as-dysfunction assumes that it does not, madness-as-strategy that it does. But this leaves the mould itself unquestioned.

By contrast, neurodiversity theorists (at least most contemporary ones) challenge this essentialist assumption, defending the need for a sociopolitical, non-essentialist analysis of cognitive and mental health categories. Contrary to common misreadings of their views by other critiques of psychiatry (see Milton and Timimi 2016) and in line with the sort of analysis proposed by many radical mad advocates (Curtis et al. 2000; P. Sedgwick 1982; see also Adler-Bolton and Vierkant 2022; Frazer-Carroll 2023), neurodiversity theorists point out the irreducibly socio-cultural and historical roots of definitions of cognitive health and normalcy, i.e., their embeddedness in particular, *contingent* social dynamics, with a special emphasis on the role of capitalist production relations, the specific human labour needs associated with it, and other intersecting social power dynamics (e.g., Chapman 2023b; Milton 2014; Walker 2021). Their proposed neurodiversity paradigm does not merely oppose the default pathologizing, "inner deficit" treatment of cognitive divergence characteristic of madness-as-dysfunction, but the *normalcy paradigm* at the root of it (Chapman and Fletcher-Watson forthcoming); one that, crucially, is also shared by traditional attempts to depathologize or "normalize" madness and divergence by forcing it into neuronormative standards of cognitive functioning.

The neurodiversity paradigm thus sees concepts of mental normalcy and mental health, as well as the association between them, not as *given*, but as reflecting contingent, and therefore contestable, sociopolitical structures and dynamics. It is this sociopolitical analysis and critique of this basic notion, cognitive normalcy, which explains what many see as a seemingly contradictory statement by neurodiversity advocates: that divergent modes of functioning may be *both* disabling, even "dysfunctional" (at least within the specific social dynamics that configure what "normal" functioning is), hence requiring the allocation of especial resources and accommodations; *and*, at the same time, worthy of respect and value, something that may ground one's identity as well as alternative notions of health and flourishing.

I think this insight is crucial for advancing Garson's own political aspirations for madness, as it reveals the limitations of the madness-as-strategy concept. For madness-as-strategy still circumscribes our ability to see value in madness within the bounds of "normalcy"; within the bounds of what we, today, perceive as a "normal" reaction to adverse life circumstances, a "natural" response of an allegedly universal cognitive architecture, a result of a pre-established, unquestioned cognitive economy that always maximizes utility—an expression of the *Homo Economicus* in the cognitive domain. But social dynamics affect us in many ways. Some forms of madness may be a completely "natural" response to them; others, however, may indeed be the result of breakdowns *precisely* caused by those dynamics. Would that sort of madness be less valuable? Questioning underlying notions of cognitive normalcy opens the door for a more radical defence of madness: one that sees value in it even when it's not the result of "everything functioning as it should"; or even *precisely because it is,* at least sometimes, the result of abnormal, disabling, dysfunctional cognition, of modes of functioning that fundamentally defy the usual order of things and its reflection in the usual assumptions concerning what a "well-oiled" mental machine is supposed to be. And yet, also precisely because of their abnormal, disabling, or dysfunctional character, mad and divergent modes of functioning may require special accommodations and resources—whether medical, psychosocial, or otherwise (Adler-Bolton and Vierkant 2022; Chapman 2023b; Frazer-Carroll 2023).

To be sure, I think the author would agree with much of this. But I nonetheless think it's crucial to stress the importance of going beyond madness-as-strategy, to tackle the sometimes-implicit assumptions about cognitive normality underpinning it. The effects of these implicit assumptions are sometimes visible throughout the book, for instance when madness is presented as actual *sanity*, i.e., as a sane response to an insane social order. Powerful as it undoubtedly is, this slogan, which appears in several parts of the book and is especially prominent in the book's most explicitly political chapter (Chapter 13), can nonetheless subtly contribute to reinforce the very standards of sanity—and the social relations underpinning them—that madness is supposed to disrupt. This is because the notion of sanity that madness is to be associated with may well still be imbued with normalcy assumptions. This is particularly evident in the 1960s counterculture's distinction between "good/true" and "bad/false" madness; between the "sane", "reprogramming", "morally awakening" madness of progressive, acidhead hippies, and the "useless", "anti-social", or "paranoid" madness of the "speed freak" punks and the shit-painting "gibbering lunatics" (see Chapter 13). Here Pete Shaughnessy and other founding members and contributors to the origins of the Mad Pride movement come to mind (see Curtis et al. 2000). They not only took pride on the good-spirited, visionary, socially valuable mad extolled by the counterculture; but also, even sometimes primarily so, on the ill-spirited, anti-social, and chaotic mad of the punk scene. And I think they might have reason to do so. Maybe it wasn't speed that "destroyed the Summer of Love" (222). Maybe the summer-of-love-madness failed to subvert the social order because it was just *too easily* assimilable within it and its concomitant ideals of cognitive normalcy; that is, because it did not fundamentally challenge it, but reasserted it in a more liberal-progressive language. Perhaps this also explains the current exploitation of its main narrative within the new microdosing-based "psychedelic renaissance" that Garson himself criticizes (219).

In sum, I think that the neurodiversity paradigm's emphasis on the relational-ecological analysis of cognitive (dis)ability, on the one hand, and its sociopolitical and non-essentialist critique of categories of mental normalcy, on the other, help us to both expand the applicability of madness-as-strategy and, at the same time, see its limitations as a tool for mad and neurodivergent liberation.

But what conceptual alternative may we develop? Although I partly agree with Garson's final reflection that it may be better to overcome the "overwhelming intellectual compulsion" to craft a new "madness-as-X" (263), I'd like to conclude with a potential alternative; not fully a concept of its own, which would in any case require more space to develop, but a new, nice catchphrase to hint at possible ways forward in the development of a more liberatory conceptual scheme: *madness-as-right*. More than dysfunction or strategy, a failure or achievement of some presumed-to-be natural design, we may think of madness in terms of social rights and entitlements: in terms of an entitlement to *disrupt*, sometimes in yet incomprehensible or unrecognizable ways, the norms that characterize current social arrangements—whether moral, aesthetic, logical, or epistemic; a right to be folk-epistemologically distasteful (see Wilkinson 2020), uninterpretable; to disrupt social dynamics that "mindshape" us into norm-conformity, intelligibility, and interpretability (McGeer 2015; Zawidzki 2024), often exerting unbearable pressure on us.

A first implication of this alternative framing is that it places value on such disruptive tendencies no matter whether they are viewed as a functional biological response or rather a breakdown in normal functioning—if there is such thing at all in the first place. To be clear, the main idea underlying this approach to madness is not new but can be found in various radical anti-capitalist approaches to mental health activism. An early example would be the Sozialistisches Patientenkollektiv (Socialist Patients' Collective), a patient-led collective formed in Heidelberg during the 1970s that revindicated the "weaponization" of illness as a revolutionary strategy against capitalist domination (see Adler and Bolton, 2022). Along these lines, Chapman (2023b) has put forward a "Neurodivergent Marxism" approach that "seeks to turn both neurodivergent disablement and illness into sites of organisation and resistance to the system that necessitates both the production and harm of both neurodivergents and neurotypicals" (146).

But viewing madness as a right or entitlement also brings another important benefit: it encourages the adoption of a thoroughly context-sensitive view of its disruptive value and prompts us to question how this entitlement is distributed across social hierarchies. Evolutionary strategies may be universally shared by all humankind; rights and entitlements aren't. Framing madness as a right allows us to ask: who has this right? And who *should* have it, but typically doesn't? Who has typically enjoyed it and who is normally dispossessed of it? Whose madness has been more often viewed through a positive lens, as virtuous, valuable, functional, the hallmark of transgressive genius and vision; and whose has been historically regarded as unvaluable, useless, suppressible, the hysteric scream in need of appease and silencing? Whose madness grants responsibility exemptions and for whom is just an added burden? It also sparks questions about how to redress these imbalances. If, after all, we can only break rules within a bedrock of rule-maintaining practices, madness' disruptive creativity necessitates sanity's grip on mundaneness. Whose madness should then, for once, recede a bit, leave some space for others' madnesses to flourish? That is, who should be encouraged to break through social conventions and norms, and who should be encouraged to merely follow suit, to leave the necessary space for such breakthrough to take place?

3. Conclusion

As stated at the beginning of the Mental Patients Union’s initial manifesto, the “The Fish Pamphlet”, madness can often be understood through an analogy with a fish caught on a hook: its impulsive, seemingly irrational attempts to escape may appear bizarre or deranged to other fish at first glance, but the meaning of these struggles becomes clear when one observes the circumstances the fish is attempting to deal with (see Mental Patients Union 1974, reproduced in Irwin et al 2000). In a later development of this manifesto and organisation, now renamed Campaign Against Psychiatric Oppression (CAPO), the authors restate the point in a slightly different manner:

> We (…) assert that “patients” are not crippled by anxiety or depression or confusion; but on the contrary they are anxious or depressed or confused because they are crippled—by circumstances over which they have little or no control, circumstances which thwart, which threaten, which confuse. When a person’s behaviour is intolerable to his/her fellow humans, it is usually because his or her situation is intolerable to him or her, and such a person may need help to change the situation they are in. (CAPO 1986, 9).

These excerpts illustrate the mad insight that Garson’s notion, madness-as-strategy, aims to further articulate and develop. But they also illustrate a more subtle point: that the language and concepts we use to approach madness are and have always been in constant evolution—from the deliberate removal of “patients” from the organization’s name to the inclusion of considerations about the “crippling” consequences of the social order. This constant, often paradoxical effort to update the language and concepts about madness reflects the challenge of articulating a conceptual framework for a political movement that aims to leave no one behind—a framework that acknowledges that madness can be as much a matter of function as it is of dysfunction or disablement, an expression of both strategy and breakdown.

This paper has sought to contribute to this task by drawing on neurodiversity ideas to examine both the strengths and limitations of madness-as-strategy as a liberatory conceptual framework for madness. Specifically, I have argued that neurodiversity theory offers both (1) a way to expand the scope of applicability of madness-as-strategy from individual to collective functioning and (2) a critique of the limitations of circumscribing the positive value and reclaimability of madness solely within the realm of functionality. Instead, I have outlined a possible way forward: to reframe madness as a matter of right or entitlement. This shift helps open up conceptual space both for reclaiming madness as an identity beyond its functional aspects and, at the same time, for raising questions about the distribution of this entitlement across social hierarchies.

As stated above, this alternative concept is not meant to provide a definitive answer to the question of how we should think about madness—this, I believe, is an inherently open-ended issue. This commitment to open-endedness, however, is what I take to be a core principle of Mad Pride: "that language can be subverted and that words derive their meanings from the contexts in which they are used" (Curtis et al. 2000, 7).

Acknowledgments

Work on this paper was fully supported by the Croatian Science Foundation, under project TIPPS, grant HRZZ-IP-2022-10-1788. I wish to thank two anonymous reviewers for providing helpful comments on a previous version of this paper.

Conflict of interest statement

One of the organisers of this book symposium, Editor-in-Chief of *EuJAP*, Marko Jurjako, is also the PI of the TIPPS project, through which my work at the University of Rijeka is funded. However, this paper is solely my own work, and to the best of my knowledge, it has undergone proper peer review in accordance with EuJAP's standards.

REFERENCES

Adler-Bolton, Beatrice, and Artie Vierkant. 2022. *Health Communism: A Surplus Manifesto*. Verso Books. https://www.versobooks.com/products/2801-health-communism.

American Psychiatric Association. 1952. *Diagnostic and Statistical Manual of Mental Disorders: DSM-I, 1st Ed*. Diagnostic and Statistical Manual of Mental Disorders: DSM-5^{TM}, 5th Ed. American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596.

———. 1967. *Diagnostic and Statistical Manual of Mental Disorders: DSM-II, 2nd Edition*. 2nd edition. The American Psychiatric Association.

———. 1980. *Diagnostic and Statistical Manual of Mental Disorders: DSM-III, 3rd Edition*. 3rd edition. The American Psychiatric Association.

Biturajac, Mia, and Marko Jurjako. 2022. "Reconsidering Harm in Psychiatric Manuals Within an Explicationist Framework." *Medicine, Health Care and Philosophy* 25: 239-49. https://doi.org/10.1007/s11019-021-10064-x.

Boorse, Christopher. 1976. "What a Theory of Mental Health Should Be." *Journal for the Theory of Social Behaviour* 6 (1): 61-84. https://doi.org/10.1111/j.1468-5914.1976.tb00359.x.

Botha, Monique, Robert Chapman, Morénike Giwa Onaiwu, Steven K Kapp, Abs Stannard Ashley, and Nick Walker. 2024. "The Neurodiversity Concept Was Developed Collectively: An Overdue Correction on the Origins of Neurodiversity Theory." *Autism* 28 (6): 1591–94. https://doi.org/10.1177/13623613241237871.

Campaign Against Psychiatric Oppression (1986). "Campaign Against Psychiatric Oppression: Introduction, Manifesto, Demands." *Asylum* 1 (1): 9.

Carel, Havi. 2023. "Vulnerabilization and De-Pathologization: Two Philosophical Suggestions." *Philosophy, Psychiatry, & Psychology* 30 (1): 73-76. https://doi.org/10.1353/ppp.2023.0013.

Chapman, Robert. "Neurodiversity Theory and Its Discontents: Autism, Schizophrenia, and the Social Model of Disability." In *The Bloomsbury Companion to Philosophy of Psychiatry*, edited by Şerife Tekin and Robyn Bluhm, 371-390. Bloomsbury Companions. London: Bloomsbury Academic, 2019. http://dx.doi.org/10.5040/9781350024090.ch-018.

———. 2021. "Neurodiversity and the Social Ecology of Mental Functions." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 16 (6): 1360–72. https://doi.org/10.1177/1745691620959833.

———. 2023a. "A Critique of Critical Psychiatry." *Philosophy, Psychiatry, & Psychology* 30 (2): 103–19.

———. 2023b. *Empire of Normality: Neurodiversity and Capitalism*. 1st edition. Pluto Press.

Chapman, Robert, and Sue Fletcher-Watson. Forthcoming. *Neurodiversity: A Very Short Introduction*. Oxford: Oxford University Press.

Cooper, Rachel. 2017. "Where's the Problem? Considering Laing and Esterson's Account of Schizophrenia, Social Models of Disability, and Extended Mental Disorder." *Theoretical Medicine and Bioethics* 38 (4): 295–305. https://doi.org/10.1007/s11017-017-9413-0.

Crompton, Catherine J, Danielle Ropar, Claire VM Evans-Williams, Emma G Flynn, and Sue Fletcher-Watson. 2020. "Autistic Peer-to-Peer Information Transfer Is Highly Effective." *Autism* 24 (7): 1704–12. https://doi.org/10.1177/1362361320919286.

Curtis, Ted, Robert Dellar, Esther Leslie, and Watson, Ben, eds. 2000. *Mad Pride: A Celebration of Mad Culture*. 2nd ed. edition. Spare Change Books.

Dwyer, Patrick. 2022. "The Neurodiversity Approach(Es): What Are They and What Do They Mean for Researchers?" *Human Development* 66 (2): 73–92. https://doi.org/10.1159/000523723.

Frazer-Carroll, Micha. 2023. *Mad World: The Politics of Mental Health*. London: Pluto Press.

Garson, Justin. 2022. *Madness: A Philosophical Exploration*. New York, NY: Oxford University Press.

Graby, Steve. 2015. "Neurodiversity: Bridging the Gap between the Disabled People's Movement and the Mental Health System Survivors' Movement?" In *Madness, Distress and the Politics of Disablement*, 231-44. Policy Press.

Hoffman, Ginger A. 2017. "Collectively Ill: A Preliminary Case That Groups Can Have Psychiatric Disorders." *Synthese* 196 (6): 2217–41. https://doi.org/10.1007/s11229-017-1379-y.

———. 2019. ""Aren't Mental Disorders Just Chemical Imbalances?," "Aren't Mental Disorders Just Brain Dysfunctions?," and Other Frequently Asked Questions about Mental Disorders." In *The Bloomsbury Companion to Philosophy of Psychiatry*, edited by Şerife Tekin and Robyn Bluhm, 59–92. Bloomsbury Companions. London: Bloomsbury Academic, 2019. http://dx.doi.org/10.5040/9781350024090.ch-004.

Irwin, Eric, Lesley Mitchell, Liz Durkin, and Brian Douieb. 2000. "The Need for a Mental Patients Union - Some Proposals." In *Mad Pride: A Celebration of Mad Culture*, edited by Ted Curtis, Robert Dellar, Esther Leslie, and Watson, Ben, 2nd ed. edition, 23–28. Spare Change Books.

Jeppsson, Sofia. 2023. "A Wide-Enough Range of 'Test Environments' for Psychiatric Disabilities". *Royal Institute of Philosophy Supplements*, 94: 39-53.

Jurgens, Alan. 2023. "Body Social Models of Disability: Examining Enactive and Ecological Approaches." *Frontiers in Psychology* 14 (March):1128772. https://doi.org/10.3389/fpsyg.2023.1128772.

Kapp, Steven K. 2020. *Autistic community and the neurodiversity movement: Stories from the frontline*. Springer Nature.

McGeer, Victoria. 2015. "Mind-Making Practices: The Social Infrastructure of Self-Knowing Agency and Responsibility."

Philosophical Explorations 18 (2): 259–81. https://doi.org/10.1080/13869795.2015.1032331.

McWade, Brigit, Damian Milton, and Peter Beresford. 2015. "Mad Studies and Neurodiversity: A Dialogue." *Disability & Society* 30 (2): 305–9. https://doi.org/10.1080/09687599.2014.1000512.

Milton, Damian. 2012. "On the Ontological Status of Autism: The 'Double Empathy Problem.'" *Disability & Society* 27 (6): 883–87. https://doi.org/10.1080/09687599.2012.710008.

———. 2014. "Embodied Sociality and the Conditioned Relativism of Dispositional Diversity." *Autonomy, the Critical Journal of Interdisciplinary Autism Studies* 1 (3): 1–7.

Milton, Damian, and Sami Timimi. 2016. "Does Autism Have an Essential Nature?" Internet publication. December 1, 2016. http://blogs.exeter.ac.uk/exploringdiagnosis/debates/debate-1/.

Nesse, Randolph M., and George Christopher Williams. 1994. *Why We Get Sick: The New Science of Darwinian Medicine*. Times Books.

Rashed, Mohammed Abouelleil. 2019. *Madness and the Demand for Recognition: A Philosophical Inquiry into Identity and Mental Health Activism*. International Perspectives in Philosophy and Psychiatry. Oxford, New York: Oxford University Press.

Rosqvist, Hanna Bertilsdotter, Nick Chown, and Anna Stenning, eds. 2020. *Neurodiversity Studies: A New Critical Paradigm*. 1st ed. Routledge. https://doi.org/10.4324/9780429322297.

Sedgwick, Jane Ann, Andrew Merwood, and Philip Asherson. 2019. "The Positive Aspects of Attention Deficit Hyperactivity Disorder: A Qualitative Investigation of Successful Adults with ADHD." *ADHD Attention Deficit and Hyperactivity Disorders* 11 (3): 241–53. https://doi.org/10.1007/s12402-018-0277-6.

Sedgwick, Peter. 1982. *Psycho Politics*. Unkant Publishers.

Shaughnessy, Pete. 2000. "Into the Deep End." In *Mad Pride: A Celebration of Mad Culture*, edited by Ted Curtis, Robert Dellar, Esther Leslie, and Watson, Ben, 2nd ed. edition, 15–22. Spare Change Books.

Shorter, Edward. 1996. *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac*. New York Weinheim.

Spandler, Helen, Jill Anderson, and Bob Sapey, eds. 2015. *Madness, Distress and the Politics of Disablement*. 1st edition. Policy Press.

Szasz, T. S. 1961. *The Myth of Mental Illness: Foundations of a Theory of Personal Conduct*. Rev. ed., Reprinted, 27. [print.]. New York: Perennial.

Wakefield, Jerome C. 1992. "The Concept of Mental Disorder." *American Psychologist*, 16.

Walker, Nick. 2021. *Neuroqueer Heresies: Notes on the Neurodiversity Paradigm, Autistic Empowerment, and Postnormal Possibilities*. Fort Worth, TX.

Wilkinson, Sam. 2020. “Expressivism About Delusion Attribution.” *European Journal of Analytic Philosophy* 16 (2): 59–77. https://doi.org/10.31820/ejap.16.2.3.

Zawidzki, Tadeusz. 2024. “Skilled Metacognitive Self-Regulation toward Interpretive Norms: A Non-Relativist Basis for the Social Constitution of Mental Health and Illness.” *Synthese* 204: 109.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 2 | 155-180
https://doi.org/10.31820/ejap.21.2.4

RECONCEPTUALIZING DELUSION: STRATEGY, DYSFUNCTION, AND EPISTEMIC INJUSTICE IN PSYCHIATRY

Eleanor Palafox-Harris[1] and Ema Sullivan Bissett[1]

[1] University of Birmingham, United Kingdom

Original scientific paper – Received: 02/09/2024 Accepted: 20/01/2025

This paper is part of a book symposium on Justin Garson's *Madness: A Philosophical Exploration* curated and edited by Elisabetta Lalumera (University of Bologna) and Marko Jurjako (University of Rijeka)

ABSTRACT

In his bold and illuminating book *Madness: A Philosophical Exploration*, Justin Garson makes a case for thinking about madness as strategy, rather than as dysfunction. The reader is invited to take away a better appreciation of the historical provenance of madness as strategy, that is, this is not a new idea, destined for the fringes or of interest only to those of a more radical bent. It is rather an idea which has firm roots in the history of psychiatry. Garson's lens is wide, he is advocating a strategy over dysfunction approach for, at least, anxiety, depression, schizophrenia (and its spectrum disorders), and delusion. In this exploratory paper, we focus on delusion. We discuss what a madness-as-strategy approach might say about delusion, and how that fits with the idea that such beliefs are evolutionarily adaptive. We turn then to explore the implications of this reconceptualization of delusion for epistemic injustice in psychiatry. Our discussions will support the idea that much of the theoretical action lies not in the distinction between dysfunction and strategy, but rather in the distinction between everyday and abnormal dysfunction.

Keywords: delusion; strategy; dysfunction; doxastic dysfunction; abnormality; epistemic injustice.

Correspondence: e.l.sullivan-bissett@bham.ac.uk

Introduction

Garson's vision is a bold one as he works through the history of psychiatry understood as containing two broad approaches to madness (dysfunction and strategy). We follow Garson in understanding the approaches in the following ways. The dysfunction approach has it that

> [W]hen someone is mad, it is because something has gone wrong inside of that person; something in the mind, or in the brain, is not working as it ought. Madness results from the breakdown of a well-ordered system; it is a defect or a dysfunction. It represents the failure of the system to achieve its natural end. (Garson 2022, 1-2)

The strategy approach on the other hand has it that

> [I]n the mad, we have a purpose being fulfilled, a movement toward a goal, a machine operating as it ought. (Garson 2022, 1)

In what follows we'll pull on the threads of this characterisation of the madness-as-strategy approach, in particular, we'll speak to the idea that some ways of being mad might be goal directed, or usefully thought of as *strategies*, without them being the outputs of evolutionary mechanisms *operating as they ought*. We focus on Garson's ideas as they might apply to delusion. Let us begin with the diagnostic criteria to get the phenomenon in sight. In the glossaries of the *DSM-IV* and the *DSM-5*, the definition of delusions is given as:

> A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person's culture or subculture (e.g., it is not an article of religious faith). When a false belief involves a value judgment, it is regarded as a delusion only when the judgment is so extreme as to defy credibility. (DSM-IV 2000, 765, DSM-5 2013, 819)

We won't go over the various ways in which this definition of delusion is contentious (for an overview see Sullivan-Bissett 2024a, 3-6). What is relevant for our purposes is that although the definition is, strictly speaking, compatible with the idea that delusions are (adaptive) strategies, it more naturally puts one in mind of delusions as outputs of dysfunction.

After all, the beliefs are described as "firmly sustained" (when, presumably, they ought to be revised in the face of "incontrovertible and obvious proof or evidence to the contrary"). It is common to focus on this feature of delusions and seek to explicate it by appeal to an abnormality in belief formation or evaluation (this is the research programme of two-factor theories of delusion). And the idea that delusions are instances of *pathology* also represents the current orthodoxy. Indeed, often this is taken as a datum, with the theoretical interest identified as providing a more precise characterisation of what the pathology looks like (see e.g. Bortolotti 2018; Miyazono 2015; Petrolini 2017, 2024; and Sakakibara 2016, cf. Bortolotti 2022; Ichino and Sullivan-Bissett 2024; Lancellotta and Bortolotti 2020).

Let us narrow our focus to *monothematic* delusions, which are those concerning a single theme which arise in otherwise healthy individuals (Coltheart, Langdon, and McKay 2007, 642). We have in mind examples of the following kind:

> Patrick has the delusion that his wife has been replaced by an imposter (Capgras delusion)
>
> Selina has the delusion that she has ceased existing (Cotard delusion)
>
> Kashmir has the delusion that Beyoncé is in love with him (erotomania)

The contents of these attitudes may strike one as sufficiently bizarre as to suggest that something is going very wrong with the subject. Together then with the diagnostic definition, the idea that delusions result from dysfunction rather than (adaptive) strategy is a natural one and is reflected in much of the contemporary research on the nature of delusion and its formation. We note this not in support of the dysfunction framework, but to give a sense of the terrain. Garson's approach then, at least in the context of delusion, marks a significant and interesting departure from much of the contemporary literature.[1] Let us first then take the terrain as the following (Figure 1):

[1] Garson primarily talks about delusions in schizophrenia, which are often polythematic and elaborated. However, we think that the idea that delusions are strategies rather than dysfunctions is more plausible for monothematic delusions, and, in any case, it is these delusions that have been discussed in the context of empiricism, a natural ally of the madness-as-strategy approach.

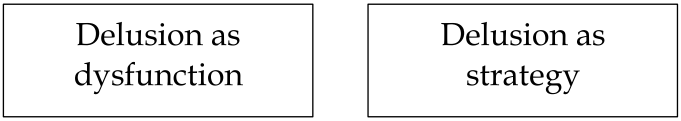


Fig. 1 Dysfunction versus strategy

1. Delusions as strategy

The idea of madness-as-dysfunction is, at least at first blush, natural, plausible, and if one is attracted to a Wakefield-style (1992) account of mental disorder, *conceptually necessary* (cf. Garson 2022, 248-9, 265). And, notwithstanding the clear strands of the madness-as-strategy approach in the history of psychiatry (for which Garson makes a magnificent case), the book, at the same time, leaves some of the finer details to the reader. Take the idea of delusions as strategy. We might well ask strategy *for what*? Garson gives some example answers from the history of psychiatry; from Griesinger (ibid., 90, 158, 184) and Freud (ibid., 4, 8) on wish fulfilment, to Johann Christian August Heinroth on delusions as coping strategies (ibid., 8), to the idea that a delusion is a "more or less deliberate diversion or escape from the unremitting tragedy of everyday life" (ibid., 125, see also 158, 169, 190), or that delusions help a person make sense of unusual feelings or experiences (ibid., 10, 226). We find this last suggestion the most promising, indeed, the idea of delusions as strategies to make sense of unusual feelings or experiences finds a home in some recent work on delusion formation.

Let us begin with predictive coding approaches (mentioned briefly by Garson, ibid., 260). According to these approaches, perceptual processing involves generating predictions about sensory input based on hypotheses about the world. Delusions are said to result from abnormalities in prediction error signalling. On what Eugenia Lancellotta has called "standard" predictive coding accounts, delusions are straightforward byproducts of a single dysfunction in prediction error minimization (2021, 56). There is no claim for delusions as strategy to be found here, let alone delusions as adaptive, indeed, Lancellotta refers to this view as the "maladaptive view of delusions" (2021, 49). However, some predictive coding theorists have conceived of delusions as generated by a *shear pin,* understood as a mechanism which is designed to break in certain conditions, to prevent further damage. A *doxastic* shear pin, that is, one at work in the machinations of belief, might be designed to break in particular circumstances (e.g. psychological distress) which would then allow for the formation of strange beliefs that would not have been formed in normal circumstances (McKay and Dennett 2009, 501-2).

Sarah Fineberg and Philip Corlett have developed an account in these terms. In the face of the anomalous prediction errors, delusions are produced which explain the anomalous experiences and allow for the resumption of learning and engagement with reality. The delusion might be thought of as a strategy resulting from shear pin breakage which enables continued learning (as against cognitive resources being used up in making sense of the experience, at the expense of continued learning and engagement with the world).

We can also find support for the idea of delusions as strategies in some empiricist approaches, which have it that delusions arise, in part, from anomalous experiences. On endorsement versions of empiricism, the delusion is an *endorsement of* what is presented in experience, and the delusional content *is identical to* the content of the anomalous experience. On explanationist versions of empiricism, the content of the anomalous experience falls short of the content taken up in delusional belief, but the delusional belief is nevertheless thought to *explain* the anomalous experience (Bayne and Pacherie 2004, 3). We'll talk in explanationist terms, although we think that both endorsement and explanationist versions of empiricism could support a delusion-as-strategy approach insofar as delusions function as explanations of experience.[2]

Let us consider some examples. A subject may have visual and auditory hallucinations of a second head on her shoulder (a case of this kind is discussed in Ames 1984). One way of making sense of this experience is to form the belief that *there is a second head on my shoulder*, after all, that's exactly the kind of experience one might expect to have if the belief were true. Or consider the anomalous experience hypothesized to precede the Capgras delusion (the belief that *someone familiar has been replaced by an imposter*). It is thought that in such cases the subject has reduced affective response to familiar faces traceable to ventromedial prefrontal cortex damage (Tranel, Damasio, and Damasio 1995). One explanation of the lack of an expected affective response to the appearance of a loved one is that *that is not in fact one's loved one, but rather someone who looks a lot like them*.

We have seen some natural theoretical allies of the idea that delusions are *strategies*. Broadly speaking, they are strategies deployed in the context of

[2] Chenwie Nie has suggested that endorsement approaches to delusion can't cast delusions as *explanations* of anomalous experiences (see Sullivan-Bissett and Noordhof 2024, 2 in reply).

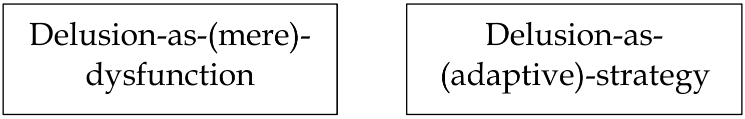
anomalous experiences which explain those experiences and keep the learning system moving. We have not yet said anything though about adaptation, something written into Garson's characterisation of madness-as-strategy. Garson is interested in the idea that delusions (among other madnesses) "could stem from mechanisms that are *performing their evolved functions perfectly well*" (Garson 2022, 251-2). Let us turn then to the idea of delusions as *adaptive* strategies.

2. Delusions as *adaptive* strategies

We are interested in the route from strategy to adaptation, between which we think there is significant water. In this section we'll explore this in the context of predictive coding and empiricism.

First, though, let us introduce Ruth Millikan's *Normal* (noting the capitalization) which will be important for some of what follows. *Normal* picks out a historical sense of normalcy rather than a statistical sense. Whilst it might be Normal for sperm to fertilize ova, it is not normal for them to do so (Millikan 1984, 34). There are Normal conditions for the proper performance of a functional item, and when a function fails to be performed, we might say that that's because the item was *dysfunctional*, or that the conditions for proper functional performance were abNormal. Now let us help ourselves to the well-worn claim that (in standard cases) mechanisms of belief production have the function of producing true beliefs (Papineau 1987; Millikan 1995). Delusions are, set against this background, clear candidates for dysfunctioning beliefs (Miyazono 2019). However, if mechanisms of belief production are operating in abNormal conditions for proper functioning, that is, in the presence of highly anomalous experiences, we have two more options for characterising them with respect to evolution. We might say that they are merely failing to perform their function due to abNormal conditions (Sullivan-Bissett 2024b), or we might say that those abNormal conditions are a trigger for a new function to be performed, and that delusions are adaptations.

One thing to note before exploring these options is that when we're thinking about delusions as responses to anomalous experience, at least in some cases, the idea that there is dysfunction *somewhere* is unavoidable. As we've noted, some experiences are hypothesized to arise from prediction errors or neurological damage, and no one wanting to argue that delusions are adaptative strategies to cope with anomalous experiences must thus deny that there's dysfunction *anywhere.* When Garson conceives of delusions as strategies to cope with anomalous experiences, he is not denying that that to which delusions are a *response* might arise from dysfunction (Garson 2022, 260). This is why we will characterise the approach in opposition to delusion-as-(adaptive)-strategy as delusion-as-(*mere)*-dysfunction (Figure 2).


Delusion-as-(mere)-dysfunction

Delusion-as-(adaptive)-strategy

Fig. 2 (Mere) dysfunction versus (adaptive) strategy

Let us turn now to whether there's a route from delusions-as-*strategy* to delusions-as-*adaptive* in predictive coding and empiricist approaches.

2.1 Predictive coding and adaptation

John Matthewson and Paul Griffiths (2017) identify four ways in which a functional trait might *go wrong*, one of which is particularly instructive here: a trait does what it is supposed to do, but its Normal conditions for doing so are ones where something else has gone wrong for the organism (2017, 454-5). This is likely the kind of picture Garson has in mind when he speculates that delusions might have "an adaptive or functional role in compensating for perceptual abnormalities, or for yielding an appearance of meaning in a seemingly absurd or cruel world" (Garson 2022, 12). Perhaps forming a belief, as against not forming a belief, is adaptive, lest one live in the paralysis of uncertainty. This idea seems to be suggested in Garson's discussion of Snyder when he says "[d]elusions are the only thing that enable me to continue to navigate my environment; the delusions rescue me from madness" (Garson 2022, 226).

Some predictive coding approaches are a natural home for thinking in these terms. Indeed, the idea that delusions are *biologically*, rather than merely *psychologically* adaptive has only been defended in this context.[3] As we have seen, some versions of the predictive coding approach focus on the importance of resuming learning and engagement with the environment. On these approaches, delusions are adaptive insofar as they maintain behavioural interactions in the face of abnormal prediction-error signaling (Fineberg and Corlett 2016; Mishara and Corlett 2009). Or, as Garson puts it, delusions could be seen as "an attempt to buffer the mind from events or perceptual experiences that would otherwise totally disrupt our ability to get around the world" (Garson 2022, 252). We see then a route from strategy to adaptation.

[3] A similar approach for delusions in schizophrenia has been defended by Pablo López-Silva (2023), although from a phenomenological framework.

We make two points here. First, as noted earlier, it is more common for predictive coding accounts of delusion to cast them as by-products of a single dysfunction (Lancellotta 2021, 56). Accounts which cast them as adaptive responses which resume learning represent deviation from these "standard" predictive coding accounts. That's no objection of course, but we might ask which of the dysfunction or adaptation versions of the predictive coding approach to delusions is to be preferred. Lancellotta (2021) has argued that the dysfunction approach is superior to the adaptation account for several reasons (including that the former is simpler and more compatible with the available empirical evidence).

In addition, some philosophers who think delusions arise from *doxastic* dysfunction have identified prediction errors as the site of such dysfunction (Miyazono 2019, 65). Recall that someone who understands delusions as adaptive strategies needn't deny that they are responses to dysfunction elsewhere (e.g. in perceptual mechanisms). However, predictive coding accounts are built upon a denial of any sharp distinction between perceptual and doxastic mechanisms. Fineberg and Corlett for example propose a single impairment in prediction error, occurring in three stages: (1) *delusional mood,* in which "attention is drawn to irrelevant stimuli", (2) *delusion formation,* in which "explanatory insight occurs and flexible processing is disabled", and (3) *explaining things with the delusion,* in which the delusion becomes habitual and "enables patients to stay engaged with the environment and exploit its regularities" (Fineberg and Corlett 2016, 4). It looks, then, like the error can characterize either perception or belief. Indeed, Fineberg and Corlett themselves note that on their model "top-down and bottom-up processes sculpt one another" (2016, 5), and Corlett and Paul Fletcher (2014) suggest that prediction error dysfunction could result in deficits in both experience and belief. So although delusions could be an adaptive strategy in the face of dysfunction elsewhere (i.e. in perceptual mechanisms), given that predictive coding approaches deny a sharp distinction between perception and belief, it is difficult to isolate the delusional response from the dysfunction which characterises the single impairment in prediction error.

All told, predictive coding accounts which take delusions to be adaptive responses to anomalous experiences break from more standard predictive coding approaches, and face challenges arising from that break regarding theoretical complexity and consistency with the empirical evidence. It might also be difficult to characterise the delusion as formed by mechanisms of belief "*performing their evolved functions perfectly well*" (Garson 2022, 251-2), given that the approach does not sharply distinguish perceptual and doxastic mechanisms.

2.2 Explanationism and adaptation

Let us return now to explanationism, which could conceive of delusions as strategies to explain anomalous experiences. From here, we won't find an easy route to the idea that these strategies are *adaptive* ones. The orthodox position in the empiricist framework is the two-factor theory, which posits (1) an experiential abnormality and (2) an abnormality in belief. A delusion may well be formed as a strategy to make sense of anomalous experience, but it is a bad strategy, and the choice of a delusional content arises from abnormalities in mechanisms of belief formation or evaluation.

Part of what motivates two-factor approaches is that delusional explanations of anomalous experiences are *bad* explanations. Indeed, Max Coltheart and colleagues suggest that any theory of delusion formation should answer two questions:

> The first question is, what brought the delusional idea to mind in the first place? The second question is, why is this idea accepted as true and adopted as a belief when the belief is typically bizarre and when so much evidence against its truth is available to the patient? (Coltheart et al. 2011, 271)

Whilst all explanationists agree that the first question is answered by appeal to anomalous experience, the second question is taken by two-factor theorists to be answerable only by appeal to a second factor (an abnormality of belief) (Davies 2009, 72). Some go further, suggesting that delusional explanations for anomalous experiences are not only *poor,* they're "unintelligible" (Nie 2023; cf. Sullivan-Bissett and Noordhof 2024, 3-5), or "nonstarters", and "the explanations of the delusional patients are nothing like explanations as we understand them" (Fine et al. 2005, 160). Delusions may well be strategies, but they are bad ones.

We are not two-factor theorists (one of us has defended the one-factor approach elsewhere, Sullivan-Bissett 2020, 2024c; Noordhof and Sullivan-Bissett 2021, 2023). In our view, the selection of the delusional hypothesis can be explained by appeal to a range of normal (which is not to say good, or rational) ways of forming beliefs that we find across human psychology, and without appeal to doxastic abnormalities. However, something being within a range of statistically normal responses does not make it adaptive, and here we can usefully distinguish between *everyday dysfunction* and *abnormal dysfunction.* This distinction allows for a defence of delusions as strategies, without abnormalities in belief, but stops short of the idea that these strategies are *adaptive*.

Doxastic dysfunction of a minor kind is likely ubiquitous. It is not exactly rare for us to fail to update background beliefs in light of new evidence, miscount, be temporarily forgetful, misuse rules of inference, and so on. We think of these as cases of *everyday dysfunctions* of belief formation. Might delusions involve dysfunctions of this kind? We think so (it would be strange if delusions were immune to the kind of everyday dysfunction that causes error in belief). Delusions involving dysfunctions of this kind would place them alongside other irrational beliefs. It is also a familiar fact that there can be motivational influences on belief; such influences are most obviously at work in cases of self-deception, but they might also be of use in explaining the formation and maintenance of delusions with welcome contents (something Garson considers on page 169).

Delusions might be conceived of as strategies, but, we say, they are strategies mounted on a host of everyday doxastic dysfunctions and, in some cases, the epistemically inappropriate influence of motivation. A better strategy when faced with the uncertainties of an anomalous experience would be to form a non-delusional belief about the origins of the experience (e.g. *I am ill,* or *I am hallucinating*). The anomalous experiences with which people with delusions must deal might represent abNormal conditions for the proper performance of belief formation. However, even on this view, things aren't going well doxastically for the subject with a delusion, and delusion formation is helped along by various errors in reasoning that might also go into explanations of other kinds of irrational beliefs. It is thus difficult to see how delusions might be conceived of as stemming "from mechanisms that are *performing their evolved functions perfectly well*" (Garson 2022, 251-2).[4]

None of this is to deny that the delusional hypothesis may hold gifts not held by the non-delusional hypothesis. And the benefits that may accrue as part of delusion's sense-making have been recently discussed in a few contexts. For example, Rosa Ritunnano and colleagues have used their case study of Harry to show that some delusions can enhance a person's sense of *meaningfulness*, understood as "the extent to which one's life is subjectively experienced as making sense, and as being motivated and directed by valued goals" (Ritunnano et al. 2022, 110). This is no doubt important, both for a more comprehensive understanding of the nature of delusion and its impact on people's lives, as well as in the clinical setting. But there's no route here to a claim of adaptation.[5]

[4] A view of this kind need not be committed to the claim that anomalous experience is *sufficient* for the formation of a delusion. Human belief formation is broad and varied, and a whole range of individual differences can contribute to the formation of a delusion in response to an anomalous experience, *or not.*

[5] This point is made by Garson in an essay for Aeon (2022).

There is also the term *epistemic innocence*, which Garson suggests is used by Lisa Bortolotti and Ema Sullivan-Bissett "to describe the apparent *reasonableness* of some delusions, given the experiences that fostered them" (Garson 2022, 260). These authors do argue that the framework of *epistemic innocence* allows for a more nuanced assessment of the epistemic and pragmatic benefits of delusions, but they also deny that the delusion is overall epistemically *good* (Bortolotti 2015, 490; Sullivan-Bissett 2018, 924)*,* and have it that delusions "compromise good functioning to a considerable extent" (Bortolotti 2015, 496).

Here then, is the theoretical terrain as we now see it (Figure 3).

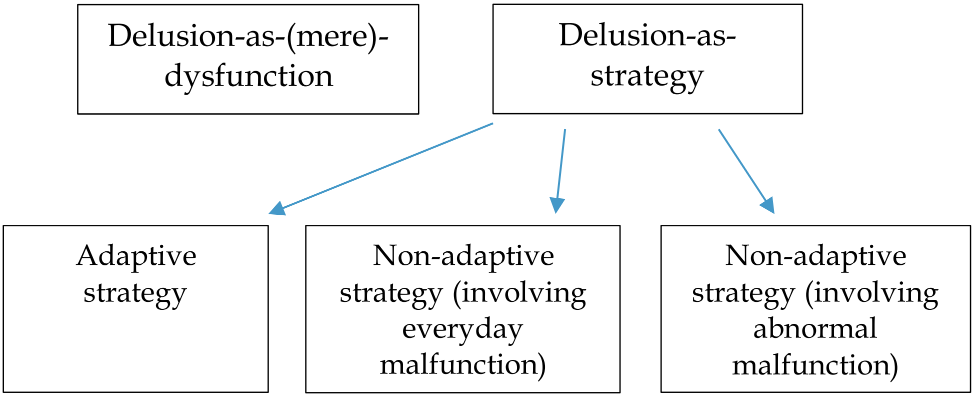

Delusion-as-(mere)-dysfunction
Delusion-as-strategy
Adaptive strategy
Non-adaptive strategy (involving everyday malfunction)
Non-adaptive strategy (involving abnormal malfunction)

Fig. 3 Delusion as dysfunction versus three strategy approaches

Overall, when it comes to understanding the formation and maintenance of delusions, our view is that the distinction between *(mere) dysfunction* and *strategy* (the top row of Figure 3) is not where the theoretical goods lie. Rather, we should explore the kind of strategy in play, and, in particular, whether the doxastic dysfunctions prompting or facilitating its execution are *adaptive,* or involve *everyday* or *abnormal dysfunction* (the bottom row of Figure 3). We think this question is more illuminating with respect to the nature and role of delusion, and it is one on which major accounts of delusion formation might be distinguished.

Of course, even if delusions arise from doxastic dysfunction of some kind, they can nevertheless be considered worthwhile epistemic contributions to interpersonal exchange, and we suspect that this is something on which we and Garson agree, but also something which is not kept in firm view in psychiatry. We turn now then to epistemic injustice and delusion. Here we think there is theoretical fruit in the distinctions of both rows of Figure 3. That is, sometimes the distinction between *(mere) dysfunction* and *strategy* might make a difference to ameliorating epistemic injustice in psychiatric encounters, whilst other times it is the distinction between *kinds of strategy* that makes the difference.

3. Delusion-as-strategy and epistemic injustice

Epistemic injustices, broadly construed, involve being wronged as an epistemic agent, in one's capacity as a *knower* (Fricker 2007, 20). Among other phenomena, this can involve having one's testimony unfairly discredited, distrusted, or discounted, as in *testimonial injustice*, having one's participation in epistemic practices unfairly undermined, as in *participatory injustice* (Hookway 2010), or having one's ability to make sense of and articulate one's own experiences unfairly constrained, as in *hermeneutical injustice*. In the literature exploring epistemic injustice and psychiatry, it is generally agreed that people with psychiatric conditions can and do experience various kinds of epistemic injustice in psychiatric contexts (e.g. Crichton, Carel and Kidd 2017; Scrutton 2017; although see Kious, Lewis and Kim 2023; Kidd, Spencer, and Harris 2023 in reply).

Epistemic injustices in psychiatry can be *interpersonal*, that is, perpetrated by healthcare practitioners or others in clinical contexts. They can also be *structural*: caused or maintained by unequal social or political dynamics. As well as these interpersonal and structural causes, Ian James Kidd and Havi Carel (2018) argue that epistemic injustices in healthcare (including psychiatry) can also be generated and exacerbated by the underlying *theoretical conception of health* which informs clinical practice and policy. Here we consider how the theoretical conception of delusion in play in psychiatric contexts might relate to the notions of *credibility*, *relevance*, and *intelligibility*, and how these affect the vulnerability of communicators with delusions to testimonial, participatory, and hermeneutical injustices.

We focus on the distinctions between delusion-as-(mere)-dysfunction and delusion-as-strategy, and between strategy accounts which posit everyday or abnormal doxastic dysfunction. The most commonly discussed cases of epistemic injustice relate to *credibility*, and here, we think the key distinction is the latter, that is, *between strategy accounts.* Nonetheless, when considering epistemic injustice related to notions of relevance and intelligibility, the broader distinction between delusion-as-(mere)-dysfunction and delusion-as-strategy can make a difference. All told, what is important for epistemic injustice in clinical encounters is not only whether we take delusions to be (mere) dysfunctions or strategies, but the details of the strategy at work.

3.1 Credibility in clinical interactions

In its most general formulation, testimonial injustice occurs when a communicator sustains an "*identity-prejudicial credibility deficit*" (Fricker 2007, 28), whereby they are perceived as having deflated credibility due to the operation of some identity prejudice. Consequently, their testimony is dismissed or distrusted. A key way in which identity prejudice deflates credibility is via stereotypes (Fricker 2007, 17).[6] Abdi Sanati and Michalis Kyratsous argue that people with delusions are frequently stereotyped as being "*bizarre*, *incomprehensible*, and *irrational*" (Sanati and Kyratsous 2015, 484), and are consequently vulnerable to testimonial injustices in clinical contexts. Here, we focus on *irrationality*, since this dimension of the stereotype relates most directly to credibility, although we touch on *bizarreness* and *incomprehensibility* in our discussion of intelligibility later (sect. 3.3).

According to Sanati and Kyratsous, the irrationality ascribed to people with delusions is not restricted to the irrational delusional belief(s), but is *generalised* to other beliefs the person holds (2015, 483). Thus, communicators with delusions are stereotyped as *globally epistemically irrational*. The irrationality that is stereotypically ascribed is not the unremarkable, everyday irrationality that even communicators without delusions exhibit. Instead, the stereotype should be understood as attributing an important difference in the *degree* or *kind* of irrationality characteristic of communicators with delusions (Palafox-Harris 2024, 261).

This stereotype deflates perceptions of testimonial credibility, as someone whose belief-system—indeed "their *general psychic life*" (Sanati and Kyratsous 2015, 484, our emphasis)—is characterised by global irrationality may not be perceived as an epistemically competent or credible communicator. If that's right, the testimony of people with delusions so stereotyped will not be given uptake in clinical interactions, and delusional communicators sustain testimonial injustices.

Let's have this basic picture in the background, as we turn to accounts of delusion and how they might impact on testimonial injustices related to assessments of *credibility*. If we think that what matters is the accuracy of the stereotype deflating credibility, then it is not (mere) dysfunction versus strategy which might make a difference here (because both kinds of

[6] For a recent discussion on the epistemic costs and benefits of stereotyping, see Puddifoot (2021) and the book symposium organized by Trakas (2025) in EuJAP.

approach, taken broadly, are consistent with the stereotype). What matters is whether the strategy involves everyday or abnormal dysfunction.

Let's start with strategy accounts which appeal to abnormal dysfunction, e.g. an account which explains the belief *that a loved one has been replaced by an imposter* by appeal to a cognitive deficit in belief evaluation. A conception of delusion that posits abnormality in belief lends theoretical justification to the credibility deficits communicators with delusions sustain in clinical interactions. Firstly, it lends theoretical support to the idea that people with delusions are characterised by irrationality that is different in kind or degree from the everyday irrationality of others. This is because positing an abnormal dysfunction in doxastic mechanisms provides a way to demarcate delusional irrational beliefs (those that *are* the product of abnormal doxastic dysfunction) from other unremarkable irrational beliefs (those that are *not* the result of abnormal doxastic dysfunction).[7] Thus, a conception of delusions as resulting from abnormality in belief reinforces the idea that people with delusions are irrational in a different way or to a greater extent than those without delusions, precisely because the irrationality associated with delusion can be traced back to an abnormal dysfunction in belief which people without delusions do not have.

Secondly, positing an abnormal doxastic dysfunction in those with delusions might also legitimise the *generalisation* of irrationality from one site (a particular delusional belief) to the presumed broader epistemic irrationality communicators with delusions are stereotypically taken to possess. It might be thought that an abnormal dysfunction in belief would lead to many of a person's beliefs being affected.[8] Conceptualising delusions as the products of abnormal doxastic dysfunction thereby sanctions the credibility deficits communicators with delusions can sustain in psychiatric contexts, because, on such a conception, people with delusions can be characterised as irrational (and thus, not credible) in a way which is global and remarkable (that is, different in an important respect from everyday irrationality). Therefore, we suggest that accounts

[7] This is not to say that a strategy account which posits abnormal dysfunction needs to be committed to delusions being the only kinds of belief that could arise from doxastic dysfunction. However, research on the etiology of other irrational beliefs (e.g. conspiratorial, self-deceptive, paranormal) has not proceeded by seeking to identify doxastic dysfunction. Rather these beliefs are theorised as arising from adaptive mechanisms, or as arising from normal range irrationalities or dysfunctions. For comparisons along these lines see Noordhof and Sullivan-Bissett (2023, 88-96) for monothematic delusion and paranormal beliefs, and Ichino and Sullivan-Bissett (2024) for monothematic delusions and beliefs in conspiracy theories.

[8] Accounts positing abnormal dysfunction can resist the stereotype of global epistemic irrationality, but this depends on the details of the second factor and delusion circumscription (Palafox-Harris 2024, 268-270).

of delusion which appeal to an abnormality in belief can contribute to testimonial injustices in psychiatry.

Let us turn to strategy accounts which deny abnormal dysfunction, e.g. an account on which the belief *that a loved one has been replaced by an imposter* can be explained as an unremarkable (which does not mean *rational*) response to weird perceptual experience, without appeal to cognitive deficit (or other abnormalities in belief). Such accounts do not support the stereotype that communicators with delusions are globally epistemically irrational, and therefore do not provide theoretical justification for downgrading the testimonial credibility of people with delusions. On an account that takes delusions as strategies for explaining anomalous experience without positing abnormal doxastic dysfunction, the irrationality of delusional beliefs is not different in kind nor in degree from other irrational beliefs. Instead, delusional beliefs can be explained by appeal to a range of *normal responses* to anomalous experience, utilizing cognitive processes which are "in no important respect different from those by which normal beliefs are formed" (Maher 1992, 262). A strategy approach of this kind does not reinforce the stereotype that people with delusions are globally epistemically irrational. If people with delusions are not irrational in a remarkable way, we avoid pre-emptively discrediting the testimony of delusional communicators on considerations of credibility.

Of course, testimonial injustices involving people with delusions might still occur in clinical interactions even if we take delusions to be strategies facilitated by everyday doxastic dysfunction. Delusional beliefs are, after all, irrational (even if they are irrational in an unremarkable way), and can contain content that is bizarre and difficult to comprehend, such as the belief *that I am dead* (a variation of Cotard delusion). Delusional testimony is therefore vulnerable to being distrusted or discounted even if we do not take delusions to result from abnormal doxastic dysfunction. Nevertheless, we suggest that if we're in the business of reducing vulnerability to testimonial injustice, it is not *(mere) dysfunction* versus *strategy* but *everyday* versus *abnormal* dysfunction within strategy approaches that could make the difference. That's because divergence on theoretical justification for the credibility deficits sustained by communicators with delusions is found *between strategy account*s, not between *(mere) dysfunction* versus *strategy*.

However, in clinical interactions, a communicator with a delusion may have her claims dismissed not only on the basis of irrationality (and therefore, of lacking credibility), but because her claims are deemed to be *irrelevant* or *unintelligible*. With respect to injustices arising from this, we think that the distinction between *(mere) dysfunction* and *strategy* is key. Let us turn to that now.

3.2 Relevance in clinical encounters

Kidd and Carel argue that:

> A conception of disease is hermeneutically influential in two related ways: it affects which experiences can be candidates for discussion and interpretation and, secondly, shapes the forms of intelligibility applicable to them. (2018, 230)

Clinicians, scaffolded by medical institutions and the theoretical framework that informs clinical practice and policy, are the arbiters of what is *relevant* in clinical encounters. It is the clinician's job to glean which information is relevant to a patient's condition from sources such as the patient's testimony, testimony from others (e.g. family members), and their medical history. Naturally, what exactly is considered relevant to clinical practice will turn on the theoretical framework for health and ill-health that is at work. If we are the papists from the Middle Ages, then the patient's moral and spiritual character—their vices, their sins—are relevant to diagnosing and treating their malady (Garson 2022, 28-29). If, however, we are Kantians, then we need only investigate which of the mind's faculties are erring (Garson 2022, 79-94).

If delusions are mere dysfunctions, the clinician's role would be to diagnose and treat the dysfunction from which the delusion arises. On such an approach, there is little room for the patient's personal narrative or interpretation. If a delusion is simply the result of a broken doxastic mechanism inside the person with the delusion, then the clinician need not look further than the dysfunction to determine how best to treat them; the patient's subjective experiences and personal interpretations sit outside the bounds of clinical relevance, and need not be given clinical uptake.

This narrow conception of relevance affects the epistemic agency of a communicator with a delusion in at least two related ways. Firstly, it constrains the *scope* or *degree* of her participation in a given clinical encounter, as the clinician alone has "special authority to delineate and treat" mental disorders (Garson 2022, 240). If the rich subjective life of the person with delusions is not clinically relevant in any substantive sense, whilst the clinical interpretation of the subject's experiences is considered authoritative, the patient's perspective need not be solicited, appealed to, nor given uptake. Indeed, the extent of the patient's participation might be exhausted by "confirming biographical details or reporting symptoms"

(Kidd and Carel 2017, 181). Thus, people with delusions might be very restricted in what areas of clinical investigation they can participate in—they can *report* symptoms, but not *interpret* them—and in the degree to which their participation is sought or taken seriously by clinicians.

Secondly, this narrow sense of clinical relevance delimits the *kind* of participation a patient can perform. According to Christopher Hookway (2010), participating in collaborative epistemic projects "is not simply a matter of exchanging information" (2010, 156). Instead, participation involves asking questions, putting forward ideas, evaluating explanations, considering possible alternative explanations, and so on. By constraining clinical relevance in a way that excludes patient contributions beyond providing information about symptoms and history, people with delusions are precluded from performing the richer participative role Hookway describes. In so doing, delusion-as-(mere)-dysfunction takes delusional communicators as sources of information for clinical inquiry, instead of as *collaborators* in the epistemic pursuits of diagnosis and treatment.

Kidd and Carel (2017) argue that "ill persons (…) are typically regarded as the objects of the epistemic practices of medicine rather than as participants in them" (2017, 181). Therefore, a (mere) dysfunction-centric view of clinical relevance treats the person with delusions as an *epistemic object* (Fricker 2007; McGlynn 2021), or perhaps as a *truncated subject* (Pohlhaus 2014), but not properly as a *participant* in clinical encounters. In restricting patient participation in these two related ways, delusional communicators sustain *participatory injustices* (Hookway 2010). Both of these effects undermine the epistemic agency of the person with delusions, whilst bolstering the epistemic authority of the clinician, thereby consolidating the *epistemic privilege* of healthcare professionals (Carel and Kidd 2014) and unequal power dynamics in clinical interaction. Thus, (mere) dysfunction accounts of delusion undermine epistemic agency by tightly circumscribing what is *relevant* to the clinical encounter, to the exclusion of the patient's rich subjective experiences.

Alternatively, if delusions are conceived of as strategies, the clinician's role is to "discern the secret purpose that madness is trying to fulfil" (Garson 2022, 1) and "target the situation or event or arrangement" (Garson 2022, 12) to which delusions are a strategic response. This conception of madness carries a much broader notion of relevance to clinical practice, and we think this broadening comes with thinking in terms of mere strategy, regardless of what kind of dysfunction the strategy involves. The personal interpretations and experiences of the communicator with a delusion are not automatically discredited as irrelevant. Rather than being relegated to the periphery of clinical investigation (if not excluded altogether), on a delusion-as-strategy framework the patient's rich subjective experiences are the very *substance* of clinical interest: in order to diagnose what the delusion is a response to, the clinician needs to look at what is happening in the patient's life, how they are experiencing their world, and how they make sense of their experiences.

This broader notion of relevance can scaffold the patient's epistemic status and agency in clinical encounters. Firstly, it means that *more* of the subject's testimony is clinically relevant compared to a (mere) dysfunction account; their personal experiences, interpretations, and meaning-making matters to a clinical inquiry that takes delusional beliefs to be *doing something*, that is, serving some purpose for the person who holds them, even if that purpose is helped along by everyday or abnormal dysfunction. In this way, the scope and degree of the person with delusion's participation is not so tightly circumscribed as on a conception of delusion as (mere) dysfunction. Secondly, conceptualising delusions as strategies allows for the patient to perform the richer participative role Hookway (2010) describes. If a patient's perspective is clinically relevant in a substantive way, then the patient is empowered to put forward their own interpretations, to participate in evaluating alternative interpretations, and to ask relevant questions. In this sense, delusional communicators are collaborators or co-investigators in the shared epistemic project of understanding their delusional beliefs, rather than the objects (or truncated subjects) of that epistemic project. Therefore, taking delusions as strategies for explaining anomalous experience can mitigate the participatory injustices people with delusions are vulnerable to in clinical encounters.

Moreover, we take the broader conception of clinical relevance contained in delusion-as-strategy to be more closely aligned than the narrow sense of relevance on delusion-as-dysfunction with frameworks like *phenomenology*, *co-production*, and *expertise-by-experience*, all of which take seriously the lived experiences of people with psychiatric conditions and promote epistemic agency. A strategy approach to delusion appears a natural ally to these more collaborative approaches to healthcare, which are often actively engaged in the project of ameliorating epistemic injustice in psychiatry (e.g. Ritunnano 2022; Carel and Kidd 2014).

3.3 Intelligibility in clinical contexts

As well as defining standards of clinical relevance, Kidd and Carel (2018, 230) emphasise that a conception of health and ill-health also defines standards of clinical intelligibility, that is, it sets the authoritative language for discussing medical conditions by producing the concepts, terms, and diagnostic categories used to interpret and talk about illness. In so doing, a conception of health and ill-health can *hermeneutically marginalise* (Fricker 2007, 153) patients. Hermeneutical marginalisation occurs when a group is not afforded equal participation in the creation of shared hermeneutical resources. Patients are hermeneutically marginalised when the interpretative resources available to make sense of and talk about an experience of ill-health are produced by healthcare professionals, rather than by people with similar experiences.

Kidd and Carel argue that aspects of patient experience "cannot gain purchase within an exclusively naturalistic conception of health", which is equipped to interpret "physiological dimensions of the process of illness" rather than subjective experiential dimensions (Kidd and Carel 2018, 228-229). A similar thing might be said for (mere) dysfunction accounts of delusion. For example, claims of *meaning* in delusion are not easily understood against a theoretical backdrop according to which delusions simply are, or arise from, dysfunction. When the theoretical conception of delusion is ill-equipped to interpret subjective experience, such experiences might either be dismissed as unintelligible, or shoehorned into the language and style the theoretical conception is equipped to interpret. Kristen Steslow argues that adopting the medicalised language used by clinicians in order to render oneself intelligible involves "forsaking the uniqueness of [one's] own perspective, understanding, and expression" (2010, 30). In this event, the subject's interpretation is given clinical uptake, but something is lost in translation. In either case, whether the patient's interpretations are dismissed as unintelligible, or made intelligible only by translation into ill-fitting medicalised concepts, the person with delusions sustains a *hermeneutical injustice* (Fricker 2007, 155): they are unfairly hindered in their own interpretative efforts by a theoretical framework that cannot accommodate such efforts.

Indeed, the problem of unintelligibility might be particularly pronounced in cases of delusion, as the lifeworld of those with delusions might be radically altered from that of the clinician, and because delusional beliefs often have bizarre content which is difficult to comprehend. As we have seen, communicators with delusions are stereotyped as not only irrational, but also bizarre and incomprehensible (Sanati and Kyratsous 2015, 484). In this way, people with delusions have to contend not only with the "communicative roadblocks" generated by an unjust hermeneutical environment which privileges the interpretative resources of clinicians (Palafox-Harris 2024, 260), but also the incommunicability of delusional experience itself.

However, conceptualising delusions as strategies to explain anomalous experience might alleviate some of the apparent incomprehensibility of delusional content. Philip Gerrans argues that "delusions often seem to be situated in chains of reasoning which, while incorrect, are *intelligible*" (2014, 113, our emphasis). A clinical approach that aims to "discern the secret purpose" (Garson 2022, 1) of a delusional belief and uncover the strategy in a delusion makes intelligible the *reasoning behind* the formation or maintenance of a delusion—reasoning which might be left obscure on an approach focused solely on diagnosing and treating dysfunctions.

Consider Capgras delusion, where someone believes *a familiar person has been replaced by an imposter*. Thinking of the Capgras delusion as a strategy for explaining anomalous perceptual experience makes the delusion more comprehensible. That is, even if the reasoning is epistemically *faulty* in some way (owing to everyday or abnormal doxastic dysfunction), we can nevertheless start to see why the particular delusional hypothesis was adopted or maintained. Understanding the strategy behind the delusion might help clinicians in understanding aspects of delusional experience and delusional content which a dysfunction-centric approach cannot make sense of. In this way, delusion-as-strategy can help to reduce the apparent incomprehensibility of delusions themselves.

Nevertheless, delusional communicators might still be hermeneutically marginalised in psychiatric contexts even on delusion-as-strategy, as we can imagine that patients would be prevented from participating *equally* in the creation of interpretative resources. However, if thinking of delusions as strategies reduces participatory injustices for the reasons we have discussed, then hermeneutical marginalisation may also be lessened. That is, a strategy approach may afford delusional communicators *more* participation in the production of shared interpretative resources than a dysfunction-centric approach. Moreover, an interpretative framework that takes delusions to be strategies for explaining or coping with anomalous experiences will be better equipped to comprehend patient interpretations of those experiences because of the standards of intelligibility such a framework carries. For example, claims that a delusional belief is *personally meaningful* are intelligible against a theoretical backdrop that assumes that delusions can serve a purpose for the person with delusions. If a particular patient interpretation is clinically intelligible, then it would not be dismissed or forced into other pre-existing concepts on the grounds of unintelligibility.

We suggest, then, that reconceptualising delusions as strategies reduces the susceptibility of delusional communicators to hermeneutical injustice in psychiatry for three reasons: (i) the standards of clinical intelligibility are equipped to comprehend patient attempts at interpretation and meaning-making, (ii) there is greater patient participation in creating hermeneutical resources, and (iii) seeking to uncover and understand the strategy behind a delusional belief reduces the apparent incomprehensibility of delusions themselves.

Of course, adopting a strategy approach to delusion (and to psychiatry generally) does not make anyone immune from acting in epistemically unjust ways. Moreover, contingent features of healthcare contexts, such as time and economic pressures, unequal power dynamics, and broader socio-political factors might still obtain even if madness-as-strategy was adopted as the underlying theoretical conception of mental ill-health. Therefore, we do not suggest that reconceptualising delusions as strategies to explain anomalous experience would automatically and wholly mitigate the various epistemic injustices delusional communicators are vulnerable to. However, strategy approaches appear better equipped to promote the epistemic agency of people with delusions than (mere) dysfunction approaches.

4. Conclusions

We have explored conceiving of delusion as (mere) dysfunctions versus strategies. We found little mileage in the idea that delusions are *adaptive* strategies, but suggested that thinking of them as strategies finds a home in some contemporary approaches to delusion formation. Garson suggests that thinking in terms of madness-as-strategy rather than madness-as-dysfunction has interesting implications for "research and treatment itself, that is, the manner in which people are healed" (Garson 2022, 11). We have suggested, instead, that the key distinction with respect to the nature of delusion is in fact *between* strategy approaches, in particular, between those that appeal to abnormal malfunction versus those that appeal to everyday malfunction. Settling this is, for example, at the heart of the debate within explanationism about the number of *factors* involved in delusion.

We have also explored implications for epistemic injustice in psychiatry, and have suggested that the distinction between (mere) dysfunction and strategy approaches to delusion matters for epistemic injustices related to notions of clinical relevance and intelligibility (participatory and hermeneutical injustices). However, what matters for testimonial injustice is whether a strategy approach appeals to abnormal or everyday dysfunction. Therefore, reconceptualising delusions as strategies (broadly conceived) rather than arising from mere dysfunction is not enough to mitigate the credibility deficits communicators with delusions sustain. To focus solely on the distinction between delusion-as-(mere)-dysfunction and delusion-as-strategy obscures important differences between strategy approaches which bear on considerations of credibility. Therefore, the alleviation or amelioration of epistemic injustices in psychiatry is best served not by a move from mere dysfunction to strategy, but rather to strategy *without abnormal doxastic dysfunction*.

In sum, when it comes to delusion, thinking in terms of *strategy* over *(mere) dysfunction* might make a difference to the amelioration of some epistemic injustices. However, overall, it is consideration of the *kind of strategy* involved in delusion which bears more theoretical fruit when considering questions of the nature of delusion and epistemic injustice in psychiatry.

Acknowledgments

We acknowledge the support of the Arts and Humanities Research Council for funding the work of which this is a part (Midlands4Cities Doctoral Training Partnership for Eleanor Palafox-Harris, and *Deluded by Experience,* grant no. AH/T013486/10 for Ema Sullivan-Bissett). Thanks also to Lisa Bortolotti and two reviewers for this journal for their comments on an earlier version.

REFERENCES

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR.* Washington, DC: American Psychiatric Association.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders. DSM-5*. Washington, DC: American Psychiatric Association.

Ames, David. 1984. “Self-shooting of a Phantom Head.” *The British Journal of Psychiatry* 145 (2): 193-4.

Bayne, Tim, and Elisabeth Pacherie. 2004. “Bottom-Up or Top-Down: Campbell’s Rationalist Account of Monothematic Delusions.” *Philosophy, Psychiatry, and Psychology* 11 (1): 1-11.

Bortolotti, Lisa. 2015. “The Epistemic Innocence of Motivated Delusions.” *Consciousness and Cognition* 33: 490-99.

Bortolotti, Lisa. 2018. “Delusion.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta, Summer 2022 edition.

Bortolotti, Lisa. 2022. “Are Delusions Pathological Beliefs?” *Asian Journal of Philosophy* 1 (31): 1-10.

Carel, Havi, and Ian J. Kidd. 2014. "Epistemic Injustice in Healthcare: A Philosophical Analysis." *Medicine, Health Care and Philosophy* 17: 529-540.

Coltheart, Max, Robyn Langdon, and Ryan McKay. 2011. "Delusional Belief." *Annual Review of Psychology* 62: 271-98.

Coltheart, Max, Robyn Landon, and Ryan McKay. 2007. "Schizophrenia and Monothematic Delusions." *Schizophrenia Bulletin* 33: 642-7.

Corlett, Philip, and Paul Fletcher. 2014. "Computational Psychiatry: A Rosetta Stone Linking the Brain to Mental Illness." *Lancet Psychiatry* 1 (5): 399-402.

Crichton, Paul, Havi Carel, and Ian Kidd. 2017. "Epistemic Injustice in Psychiatry." *BJPsych Bulletin* 41 (2): 65-70.

Davies, Martin 2009: "Delusion and Motivationally Biased Belief. Self-Deception in the Two-Factor Framework." In *Delusion and Self-Deception*, edited by Tim Bayne and Jordi Fernández, Jordi, 71–86. Psychology Press.

Fine, Cordelia, Jillian Craigie, and Ian Gold. 2005. "The Explanation Approach to Delusion." *Philosophy, Psychiatry, and Psychology.* 12 (2): 159-63.

Fineberg, Sarah K., and Philip R. Corlett. 2016. "The Doxastic Shear Pin: Delusions as Errors of Learning and Memory." *Cognitive Neuropsychiatry* 21 (1): 73-89.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Garson, Justin. 2022. *Madness: A Philosophical Exploration.* New York: Oxford University Press.

Garson, Justin. 2022. "The Helpful Delusion." *Aeon. https://aeon.co/essays/evidence-grows-that-mental-illness-is-more-than-dysfunction*

Gerrans, Philip. 2014. *The Measure of Madness: Philosophy of Mind, Cognitive Neuroscience, and Delusional Thought*. Cambridge, MA: MIT Press.

Hookway, Christopher. 2010. "Some Varieties of Epistemic Injustice: Reflections on Fricker." *Episteme* 7 (2): 151-163.

Ichino, Anna, and Ema Sullivan-Bissett. 2024. "Conspiracy Beliefs and Monothematic Delusions: A Case for De-patholologizing." *Erkenntnis.* https://doi.org/10.1007/s10670-024-00881-w

Kidd, Ian J., and Havi Carel. 2017. "Epistemic Injustice and Illness." *Journal of Applied Philosophy* 34 (2): 172-190.

Kidd, Ian J., and Havi Carel. 2018. "Healthcare Practice, Epistemic Injustice, and Naturalism." In *Harms and Wrongs in Epistemic Practice*, edited by Simon Barker, Charlie Crerar, and Trystan S. Goetz, *Royal Institute of Philosophy Supplements:* 84, 211-233. Cambridge: Cambridge University Press.

Kidd, Ian J., Lucienne Spencer, and Eleanor Harris. 2023. "Epistemic Injustice Should Matter to Psychiatrists." *Philosophy of Medicine* 4 (1): 1-4.

Kious, Brent M., Benjamin R. Lewis, and Scott Y. H. Kim. 2023. "Epistemic Injustice and the Psychiatrist." *Psychological Medicine* 53 (1): 1-5.

Lancellotta, Eugenia. 2021. "Is the Biological Adaptiveness of Delusions Doomed?" *Review of Philosophy and Psychology* 13: 47–63.

Lancellotta, Eugenia, and Lisa Bortolotti. 2020. "Delusions in the Two-factor Theory: Pathological or Adaptive?" *European Journal of Analytic Philosophy* 16 (2): 37-57.

López-Silva, Pablo. 2023. "Minimal Biological Adaptiveness and the Phenomenology of Delusions in Schizophrenia." In *The Philosophy and Psychology of Delusions: Historical and Contemporary Perspectives* edited by Ana Falcato and Jorge Gonçalves, 126-140. New York: Routledge.

Maher, Brendan. 1992. "Delusions: Contemporary Etiological Hypotheses." *Psychiatric Annals* 22 (5): 260-8.

Matthewson, John, and Paul Griffiths. 2017. "Biological Criteria for Disease: Four Ways of Going Wrong." *Journal of Medicine and Philosophy* 42: 447-66.

McGlynn, Aidan. 2021. "Epistemic Objectification as the Primary Harm of Testimonial Injustice." *Episteme* 18 (2): 160-176.

McKay, Ryan, and Daniel Dennett. 2009. "The Evolution of Misbelief." *Behavioral and Brain Sciences* 32: 493-561.

Millikan, Ruth. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

Millikan, Ruth. 1995. "Explanation in Biopsychology." In *White Queen Psychology and Other Essays for Alice*, authored by Ruth Millikan, 171-192. Cambridge, MA: MIT Press.

Mishara, Aaron L., and Philip Corlett. 2009. "Are Delusions Biologically Adaptive? Salvaging the Doxastic Shear Pin." *Behavioral and Brain Sciences* 32 (6): 530-1.

Miyazono, Kengo. 2015. "Delusions as Harmful Malfunctioning Beliefs." *Consciousness and Cognition* 33: 561-73.

Miyazono, Kengo. 2019. *Delusions and Beliefs.* London and New York: Routledge.

Nie, Chenwei. 2023. "Revising Maher's One-Factor Theory of Delusion." *Neuroethics* 16 (15): 1-16.

Noordhof, Paul, and Ema Sullivan-Bissett. 2021. "The Clinical Significance of Anomalous Experience in the Explanation of Monothematic Delusions." *Synthese* 199: 10277-10309.

Noordhof, Paul, and Ema Sullivan-Bissett. 2023. "The Everyday Irrationality of Monothematic Delusion." In *Advances in*

Experimental Philosophy of Action, edited by Paul Henne and Sam Murray, 87-111. Bloomsbury.

Palafox-Harris, Eleanor. 2024. “Delusion and Epistemic Injustice.” In *The Routledge Handbook of Philosophy of Delusion*, edited by Ema Sullivan-Bissett, 258-273. Oxon: Routledge.

Papineau, David. 1987. *Reality and Representation*. Oxford: Basil Blackwell Limited.

Petrolini, Valentina. 2017. “What Makes Delusions Pathological?” *Philosophical Psychology* 30 (4): 502-23.

Petrolini, Valentina. 2024. “Delusion and Pathology.” In *The Routledge Handbook of Philosophy of Delusion*, edited by Ema Sullivan-Bissett, 33–45. Oxon: Routledge.

Pohlhaus, Gaire, Jr. 2014. “Discerning the Primary Harm in Cases of Testimonial Injustice.” *Social Epistemology* 28 (2): 99-114.

Puddifoot, Katherine. 2021. *How Stereotypes Deceive Us*. Oxford: Oxford University Press.

Ritunnano, Rosa. 2022. “Overcoming Hermeneutical Injustice in Mental Health: A Role for Critical Phenomenology.” *Journal of the British Society for Phenomenology* 53 (3): 243-260.

Ritunnano, Rosa, Clara Humpston, and Matthew R. Broome. 2022. “Finding Order Within the Disorder: A Case Study Exploring the Meaningfulness of Delusions.” *BJPsych Bulletin* 46: 109-15.

Sakakibara, Eisuke. 2016. “Irrationality and Pathology of Beliefs.” *Neuroethics* 9: 147-57.

Sanati, Abdi, and Michalis Kyratsous. 2015. “Epistemic Injustice in Assessment of Delusions.” *Journal of Evaluation of Clinical Practice* 21 (3): 479-485.

Scrutton, Anastasia. 2017. “Epistemic Injustice and Mental Illness.” In *The Routledge Handbook of Epistemic Injustice*, edited by Ian J. Kidd, José Medina, and Gaile Pohlhaus Jr., 347–355. Abingdon: Routledge.

Steslow, Kristen. 2010. “Metaphors in Our Mouths: The Silencing of the Psychiatric Patient.” *Hastings Centre Report* 40 (4): 30-33.

Sullivan-Bissett, Ema. 2018. Monothematic Delusion: A Case of Innocence from Experience.” *Philosophical Psychology* 31 (6): 920-47.

Sullivan-Bissett, Ema. 2020. “Unimpaired Abduction to Alien Abduction: Lessons on Delusion Formation.” *Philosophical Psychology* 33 (5): 679-704.

Sullivan-Bissett, Ema 2024a. “Introduction.” In *The Routledge Handbook of Philosophy of Delusion*, edited by Ema Sullivan-Bissett, 1-29. Oxon: Routledge.

Sullivan-Bissett, Ema. 2024b. “Monothematic Delusions are Misfunctioning Beliefs.” *Synthese* 204, 157. https://doi.org/10.1007/s11229-024-04803-9.

Sullivan-Bissett, Ema. 2024c. “The One-factor Theory.” In *The Routledge Handbook of Philosophy of Delusion*, edited by Ema Sullivan-Bissett, 414–29. Oxon: Routledge.

Sullivan-Bissett, Ema and Paul Noordhof. 2024. “Revisiting Maher’s One-Factor Theory of Delusion, Again.” *Neuroethics* 17 (1): 1-8.

Trakas, Marina. 2025. “Stereotypes Deceive Us, but Not in the Way We Commonly Think: Introduction to the Book Symposium.” *European Journal of Analytic Philosophy* 21 (1): 1-8. https://hrcak.srce.hr/328465

Tranel, Daniel, Hanna Damasio, and Antonio R. Damasio. 1995. “Double Dissociation between Overt and Covert Face Recognition.” *Journal of Cognitive Neuroscience* 7 (4): 425-32.

Wakefield, Jerome C. 1992. “The Concept of Mental Disorder: On the Boundary between biological Facts and Social Values.” *American Psychologist* 47: 373-88.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 2 | 181-204
https://doi.org/10.31820/ejap.21.2.5

MADNESS REVISITED: REPLIES TO CONTRIBUTORS

Justin Garson[1]

[1] Hunter College and The Graduate Center, City University of New York, USA

Discussion – Received: 29/04/2025 Accepted: 28/05/2025

This paper is part of a book symposium on Justin Garson's *Madness: A Philosophical Exploration* curated and edited by Elisabetta Lalumera (University of Bologna) and Marko Jurjako (University of Rijeka)

ABSTRACT

The following provides the author's responses to the four commentaries on *Madness: A Philosophical Exploration*, written by Muhammad Ali Khalidi, Eleanor Palafox-Harris and Ema Sullivan-Bissett, Miguel Núñez de Prado Gordillo, and Sofia Jeppsson and Paul Lodge.

Keywords: mental disorder; natural kind; madness-as-dysfunction; madness-as-strategy; psychosis; delusion; genealogy.

Correspondence: jgarson@hunter.cuny.edu

1. Response to Khalidi

I'm grateful for Khalidi's penetrating response, which ranges over questions about the nature of conceptual genealogy, the political dimensions of madness, and the future of psychiatry as a unified discipline. I cannot possibly do justice to the richness of his suggestions, but I will say a brief word about each.

1.1 Madness as conceptual genealogy

At the outset of my book, I rejected the idea that *Madness* is an exercise in either the history of science or "conceptual genealogy", by which I meant a project that purports to answer empirical-historical questions (e.g., "When did madness-as-dysfunction originate? How did it come to dominate mental health thinking today?"). The primary goal of my book is not to provide a causal explanation, nor is it to explain how the dysfunction-centered perspective has become so entrenched in today's mental health thinking and practice. It is an attempt to use historical texts to construct new concepts: *madness-as-dysfunction* and *madness-as-strategy*. Of course, once these concepts are available, a historian may wish to use them to frame a proper historical narrative.

Yet Khalidi argues that, understood in a specific way—namely, in the way Foucault (1977) characterizes genealogy—my book is an exemplar of genealogy. I agree entirely with Khalidi that I was too hasty in rejecting the idea that the book may be a genealogy in some *sense* of that term. In rejecting that my book is a "genealogy", I wanted to distance myself from a certain mode of philosophizing that is primarily causal-explanatory in its intent, in a way that is perhaps closer to what Queloz (2021) calls "conceptual reverse-engineering". This is the sense of genealogy that I had taken from Nietzsche's 1877 *On the Genealogy of Morals*. A conventional way of understanding Nietzsche's work is that it sketches a causal-historical explanation for the existence of a distinctive (and, he thinks, distinctively perverse) value system that, roughly, equates "good" with weakness and "evil" with strength. For Nietzsche, the existence of this value system requires a historical account, much like how one might give a historical account of World War II, or the Irish potato blight of 1845.

I agree with Khalidi that *if* we construe genealogy from the standpoint that Foucault describes in 1977—as, in a sense, the precise *antithesis* of a historical account—as an attempt to provide not an origin story, but to disrupt the possibility of origin stories, to find multiplicity rather than singularity, and to demonstrate discontinuity rather than continuity—then my work comes much closer to genealogy.

Yet there is one aspect of Foucault's characterization of genealogy that I want to resist. In some ways, my goal is the opposite of the one Foucault describes. If Foucault wants us to dissolve the unity of an origin into a multiplicity of competing goals, desires and agendas, I, in contrast, want to retrieve a certain conceptual unity from that apparent multiplicity. I want to emphasize the extent to which these texts, produced in different eras and motivated by different worldviews, form a surprising unity, "a secret resonance, a *spiritual brotherhood*" (Garson 2022a, 3), but one that is not connected by chains of influence. Whether we think of madness as a divine punishment designed to redeem fallen humanity, or as a compromise between conscious and unconscious desires, or as the working out of the *vis medicatrix naturae*, or an evolved adaptation, we are seeing teleology rather than dysteleology. It is in this sense that I think the modern evolutionary theory that depression is a designed signal is far closer to Robert Burton's theological view of melancholy as a divine intervention than it is to the chemical imbalance theory of depression of the 1980s and 1990s. The reason it is difficult to see the conceptual continuity between these various texts is because we simply lack the concepts to.

1.2 Fanon and the biopsychosocial paradigm

I now turn to Khalidi's invocation of Fanon, which aligns neatly with the social and political aims of my work. What I want to highlight is Khalidi's insightful observation that for Fanon, madness does not merely have a political *cause* (even the proponent of madness-as-dysfunction will happily acknowledge as much), but more deeply, it can arise as a strategic *response* to that political situation. Khalidi helpfully points out that for Fanon, madness can be adaptive, functional, and purposeful, rather than as a pathology that happens to be "triggered" by external forces.

The mere idea that madness has a social cause (as Khalidi notes) has indeed been a part of traditional discourse, even reaching back before Fanon. My favorite example comes from George Cheyne, whose *The English Malady* of 1818 depicts melancholy as a response to British imperialism and, strangely enough, to the international spice trade. As I summarize his view

> The reason for the prevalence of melancholy amongst the English is their damnable need to import heavy foods, rich wines, and abundant sauces and spices from the four corners of the earth, imbibements that God never intended the English to enjoy. (Garson 2022a, 7)

God uses melancholy as a signal "to draw our attention to the rebelliousness of our hearts" (Ibid.). This idea that madness has social causes is also alluded to, as Khalidi notes, in the more contemporary notion of a "biopsychosocial" paradigm (Engel 1977).

Yet, I think Fanon is doing something quite distinctive in the texts at issue, and it is a virtue of Khalidi's discussion to draw it out. There are two very different claims that need to be separated when we say, as many do, that we need to better "understand the social causes of mental illness". I think by confusing these two claims under the rubric of the "biopsychosocial" view, it has become too easy for psychiatrists and other mental health professionals to present themselves as being far more progressive and politically attuned in making sense of mental health problems than they really are.

The first claim is that social factors, such as poverty, migration, or climate change, can initiate a causal sequence that culminates in some kind of internal pathology (e.g., Jester et al. 2023). This is still an instance of madness-as-dysfunction, despite the fact that seeks the ultimate causes of madness in the social world. (In the context of the French occupation of Algeria, one might imagine a doctor saying, "It is so unfortunate that colonialism has broken the minds of young Algerians, making them hallucinatory, delusional, and aggressive.") The problem is that this seemingly inclusive viewpoint doesn't succeed in moving beyond the internalizing, pathologizing tendency of the biomedical framework as whole (Read 2005).

The second claim is that the symptoms of disorder—hallucinations, delusions, aggressiveness, paranoia, anxiety, and suicidal ideation—can be, in a sense, *strategic responses* to political situations. These are survival strategies, not brain defects—much less brain defects "triggered" by social causes. It seems to me that this is precisely what distinguishes Fanon's vision from modern invocations of the "biopsychosocial" model. As Khalidi observes

> [Fanon] turns the tables on colonial psychiatry by asserting that the alleged laziness and intransigence of natives under colonial domination are in fact not pathologies at all, but the natural state of resistance to colonialism. (Khalidi this issue, 14)

What Khalidi articulates is not merely a recognition of the social dimension of mental health problems but an acknowledgment that these problems may be functional responses to the social situation, or, as Laing memorably put it, a "sane response to an insane world". This is where we gain the full benefit of moving away from the individual pathologizing perspective toward sociocultural critique. You could say that both perspectives draw attention to social causation, but one still maintains a pathologizing tendency that the other abandons.

1.3 The (dis)unity of mental illness

I have less to say about Khalidi's provocative suggestion that if we accept that some forms of madness (or, equivalently, "psychiatric condition", "mental illness", "mental disorder") are in some sense "by design", then that would fragment the very concept of madness. After all, many theorists, such as Boorse and Wakefield, have argued that the chief factor that unifies all of the diverse conditions we characterize as mental illness is that they stem from an underlying dysfunction, or a failure of something in the individual to "do what it is supposed to do". This consideration leads Khalidi to suggest that, if some forms of madness are truly by design, we appear to abolish the very category of madness ("mental illness", "mental disorder", etc.) as a natural kind or even as a unified field of study. That, in turn, would raise a challenge to the very status of psychiatry as a branch of medicine: if mental illness is not a single kind, but a hodge-podge of conditions marked by vague intuitions about suffering, social deviance, or abnormality, then psychiatry has no proper object. (By the same token, there's no such thing as a science of weeds.) And without a proper object, critics have no basis for reprimanding psychiatry for "overreach". Psychiatry could neither succeed nor fail to stay within its bounds.

It's worth noting that, in some ways, these provocative implications echo those raised by Thomas Szasz (1960). One way of reading his complex article is to see it as arguing that mental illness does not exist as a natural kind. Instead, "mental illness" is similar to "jade", which refers to the disjunctive kind *jadeite-or-nephrite*. For Szasz, "mental illness", similarly, lumps together two quite different categories. There are brain disorders, such as Alzheimer's, which fall under the purview of neurology, and there are problems of living, which fall under the purview of psychotherapy. Lacking a unified domain, psychiatry dissolves.

Khalidi argues that we can rescue psychiatry from this unhappy fate. He points out that there are other domains of medicine that study diverse sets of phenomena, such as pediatrics. Yet perhaps I betray my Szaszian leanings by *welcoming* the conclusion that perhaps psychiatry should dissolve and be replaced by an array of specific mental health practices and interventions, such as peer support, psychotherapy, neurology, social work, and so on. If, as I argue (Garson forthcoming), psychiatry is *essentially* wedded to a dysfunction-centered framework—that is, if part of what makes psychiatry, psychiatry, is that it depicts mental health challenges as symptoms of disorders—then perhaps the needs of troubled people would be better supported by alternative interventions.

2. Response to Palafox-Harris and Sullivan-Bissett[1]

2.1 Points of convergence

I want to applaud Palafox-Harris and Sullivan-Bissett for wading through the complex morass of terminology and concepts that have arisen regarding the understanding of function, dysfunction, and delusion. Not only do we have the distinction between madness-as-dysfunction and madness-as-strategy, but *within* strategy models, we also encounter the important distinctions between adaptation (evolutionary vs. ontogenetic, which can clash), adaptive (psychological vs. biological, which can also clash), and hybrid dysfunction/strategy models (such as a standard explanation for Capgras syndrome as involving a "functional" response to a perceptual "dysfunction"). To this complex landscape, Palafox-Harris and Sullivan-Bissett add a new distinction between "abnormal dysfunction" and "everyday dysfunction".

The conceptual and terminological difficulties multiply when we consider the range of models of delusion, including predictive coding frameworks (of both dysfunction and hybrid dysfunction/strategy varieties), explanationist approaches (including one-factor and two-factor theories), traditional psychodynamic approaches, and phenomenological approaches. It's not always clear how one ought to situate these theories in relation to one another.

Given the conceptual and empirical complexities we are navigating, I want to take a moment to step back and summarize the points of agreement between myself and Palafox-Harris and Sullivan-Bissett. This is crucial because it highlights what we all take to be the most important problem: how to ethically engage with those whose beliefs are labeled "delusional", or, more tellingly, who are simply labeled "delusional" as persons.

I see three key points of agreement between Palafox-Harris and Sullivan-Bissett and myself. First, we agree that the paradigm I describe as madness-as-dysfunction, and what they describe as the "mere dysfunction" model, is deeply entrenched in current mental health orthodoxy. Despite the fact that for over two centuries, clinicians have regularly observed that delusions might, in some ways, be beneficial or even "designed",

[1] I'm grateful to Pablo López-Silva for thoughtful feedback on this section.

psychiatry, and the mental health professions more generally, remain dominated by the dysfunction model. While there are a handful of voices emphasizing "strategy" approaches to delusions, including Fineberg and Corlett (2016), Sullivan-Bissett (2022), Isham et al. (2022), Ritunnano et al. (2022), López-Silva (2023), Bortolotti (2023), and Garson (2024a), they remain in the slim minority. When it comes to the issue of theory evaluation in the sciences, it is important not to pretend that strategy and dysfunction approaches to delusions are on a "level playing field".

Secondly, we agree that delusions are functional in *some* clinically meaningful sense, and that this fact must be central to mental health practice. (What are the delusions "doing" for the patient? How can we take this "function" into account when crafting a treatment plan?) What we primarily disagree about is whether delusions are functional in the sense of being adaptations—a matter of biological design, or as I put it, the output of "mechanisms that are performing their evolved functions perfectly well". They think the evidence against the adaptationist view is far stronger than I think it is.

Third, despite these conceptual and empirical puzzles of delusion, our primary mission is a moral one: how can we best avoid the systematic epistemic injustice that people labeled "delusional" are routinely subjected to? We must not only conceptualize delusions in a way that is empirically accurate, but also deeply respectful of the agency, humanity, and reasonableness of people who are ordinarily dismissed as "irrational", "crazy", or "not worth listening to" (Garson 2024c). In this connection, it's noteworthy that in these scholarly discussions, such as the one we are engaged in here, people with delusions are generally the *them* (the object under discussion) in contrast to the *us* (we who are presumptively non-delusional, and thus possess enough sense to talk cogently about them).

My aims here are limited: I simply want to elaborate on my claim that delusions "could stem from mechanisms that are, in fact, *performing their evolved functions perfectly well*". Despite their objections, I continue to think that, when properly articulated, this remains a viable starting point for approaching delusions. To properly articulate it, though, I must sketch the underlying theory of biological function that my work is rooted in. Because of space limitations, I won't address Palafox-Harris and Sullivan-Bissett's illuminating discussion of epistemic justice, except to note that I strongly agree that theorists have a deep ethical imperative to minimize epistemic injustice in all of its forms.

2.2 The prediction error model

Palafox-Harris and Sullivan-Bissett agree that there is some limited sense in which we can see delusions as “strategic”, but they believe we can acknowledge this without going so far as to say that delusions are a matter of biological design. Put differently, they wish to problematize the inference from strategy to adaptation. I agree that there is no easy inference from “delusions may benefit people in some way or another” to “delusions are adaptations, specifically shaped by natural selection to deliver a benefit”, or, even more weakly (and closer to my own view), that “delusions are the outputs of a cognitive mechanism functioning exactly as designed”. Adaptationist inferences are always risky in multiple ways (Garson 2022b, Chapters 3 and 4). But I don’t see any compelling theoretical or empirical reasons to take the adaptationist view off the table—for example, as suggested by Lancellotta (2022) in an essay entitled “Is the biological adaptiveness of delusions doomed?”

To survey the problems with adaptationist views, Palafox-Harris and Sullivan-Bissett consider two different approaches to delusion, the prediction error approach, and the explanationist approach. (While some might see the prediction error approach as a form of explanationism, they keep the two approaches separate.) The prediction error approach holds that delusions stem from some disruption to the “prediction error” signal, the signal that’s supposed to tell us whether our perceptual inputs conform to our top-down expectations about the way the world should be. The specific version of the prediction error approach endorsed by Fineberg and Corlett (2016), however, is quite complex. It depicts delusions *both* as a product of a dysfunctional prediction error, and *also* as performing a function—namely, the function of helping the individual conserve cognitive resources in a profoundly disorienting situation. In this way, their view embodies an intriguing mixture of dysfunction and strategy approaches.

Lancellotta (2022) has objected to their hybrid function-dysfunction model on the grounds that it is unnecessarily complex relative to mere dysfunction models, and Palafox-Harris and Sullivan-Bissett agree with her assessment. In Lancellotta’s view, a theory that accounts for delusions merely in terms of a dysfunction is simpler, and therefore preferable, to one that depicts delusions both in terms of a dysfunction and a function.

I found Lancellotta’s argument less than persuasive, for three reasons. First, Fineberg and Corlett are trying to reconcile over two centuries of clinical data (see references above) as well as their own empirical research, that shows that delusions benefit people in certain ways, and that these benefits seem to play some role in their "fixity", that is, their resistance to counterevidence. Their hybrid dysfunction/strategy model is one way to reconcile this complex data. To get a parsimony argument off the ground, Lancelotta would have to show that the "mere dysfunction" model explains that data equally well, which I don't think she does. (Incidentally, parsimony arguments are notoriously difficult to apply—see Sober and Wilson 1998, 291-295. Researchers who hold competing theories rarely agree on which is the simpler theory, which data those competing theories are meant to account for, and how to spell out the tacit *ceteris paribus* conditions.)

Second, there is nothing particularly surprising about the idea that the mind is replete with mechanisms that are, to put it colorfully, *designed to fail.* This is, of course, the chief theoretical innovation of Freud's *The Psychopathology of Everyday Life.* There, Freud points to a wide range of such "designed failures": forgetting, misplacing items, slips of the tongue. He calls all of them aptly *Fehlleistung*, a term that his biographer Ernest Jones unfortunately translated as "parapraxis", but would be more accurately translated as "faulty performance". Forgetting the name of an ex-lover can be seen as a minor breakdown of memory—after all, the whole *purpose* of memory is to retrieve facts and present them to consciousness. Yet, in another sense, it's a *designed* breakdown. It's a way of managing anxiety or grief. Self-deception, as Trivers (2011) famously argued, is another example of such a designed failure. True, it pays, evolutionarily, to have an accurate representation of your own abilities (for example, if you're wandering into combat unprepared). But it also pays, evolutionarily, to have an *inflated* sense of your own capacities (for example, when it comes to taking on challenges that you would have otherwise avoided). There's nothing ontologically profligate about seeing delusions as one more example of a "designed failure".

Finally, I worry that Lancellotta overstates the distinction between psychological adaptiveness and biological adaptiveness. She points out correctly that even if delusions can improve one's self-image, that doesn't mean that they have fitness advantages that can be cashed out in terms of survival or reproduction—the only criterion that evolution really cares about. But on this score, I agree entirely with López-Silva's (2023) recent observation that disentangling biological and psychological adaptiveness is actually quite difficult. This is because, as he points out (echoing Fineberg and Corlett 2016), delusions in the context of schizophrenia are often preceded by a "prodromal" period marked by an intense sense of disorientation. The delusion doesn't just explain something that I'd otherwise find puzzling; it brings a kind of stability into my world that enables me to function well enough to meet my basic survival needs.

There is a second reason that it's quite difficult to disentangle biological and psychological adaptiveness: one benefit of achieving psychological well-being, quite generally (that is, even outside the context of mental illness) is that it makes us less likely to commit suicide. Swanepoel and Soper (2025), building on Soper (2018), recently argued that suicide is a much more serious danger for our species than we typically realize. They think that any organism that possesses both the capacity to suffer, and the cognitive capacity to understand its own mortality, would consider hastening its own death as a solution to suffering. Consequently, they conjecture that evolution likely equipped us with a host of "anti-suicide" devices—cognitive quirks that prevent us from taking our own lives. Delusions, they conjecture, are just one of those designed "quirks". By giving us a sense of understanding, self-worth, or purpose, they argue, delusions literally give us a reason for living.

2.3 The explanationist model

In explanationist models, delusions are considered to be explanations for anomalous perceptual experiences. Capgras delusion, for example, describes the belief that a loved one has been replaced by a perfect imposter. One prominent theory of Capgras is that there's a dysfunctional communication failure between the perceptual and affective parts of my brain. That leads to an absence of an emotional response upon seeing my loved one. The patient then begins looking for explanations for this disorienting experience. The explanation that they arrive at: *my loved one must've been replaced by a perfect imposter.*

Explanationist models split into two-factor approaches, which depict delusions as resulting from *both* a perceptual dysfunction and a cognitive dysfunction, and one-factor approaches, which depict delusions as the output of a cognitive mechanism functioning fairly "normally" in the face of a perceptual dysfunction. Put differently, according to the one-factor approach, we don't need to postulate any distinctive cognitive malfunction in order to understand why somebody would arrive at such an unusual belief. I had one-factor theories in mind when I wrote that delusions "could stem from [cognitive] mechanisms that are *performing their evolved functions perfectly well* [when confronted with such unusual perceptual data]".

Palafox-Harris and Sullivan-Bissett still object to my way of putting things. Clearly, the delusional mind is *not* working "perfectly well", cognitively speaking. If it *were* working perfectly well, then when faced with this perceptual anomaly, you would imagine the sufferer thinking something like this: *That's funny—I suddenly feel perfectly cold toward my* *wife. Perhaps there's something wrong with my brain. Or perhaps I'm more ambivalent about our relationship than I was willing to acknowledge. Or perhaps she's not actually my wife but a cleverly-designed impostor. At any rate, I should probably go see a neurologist or therapist before exploring this rather far-fetched theory.*

I agree that such an inner monologue would be more epistemically *ideal*. But there's a sharp distinction between the norms that govern our epistemic ideals, and the norms that govern (biologically) proper functioning. Perhaps the reasoning powers in the Capgras sufferer are less than "ideal" in some human sense. But that doesn't mean they're not functioning exactly as designed by evolution. In fact, I think one of the most striking findings of modern evolutionary theorizing, particularly as its embodied in the movement known as Darwinian medicine (Gluckman et al. 2009), and equally, the movement known as evolutionary psychiatry (Abed and St. John-Smith 2022), mechanisms operating "perfectly well" according to evolved norms of proper functioning can fall far short of what we might consider "ideal" in the modern world: a canonical example is our human proclivity toward sugary and fatty foods. A standard explanation for this contemporary ailment is that it represents the outcome of cognitive mechanisms functioning exactly as they are (evolutionarily) meant to despite the fact that, given that our modern environments are replete with such foods, that proclivity leads to suboptimal health outcomes.

For a more extreme and unusual example, consider the case of "imprinting gone awry", a thought experiment devised by Jerome Wakefield (1999), which Fagerberg and Garson (2024) discuss at length. At birth, goslings have an imprinting mechanism—a neural device that's meant to "imprint" upon the first large moving object the gosling sees. In the (statistically) ordinary case, the first large moving object the gosling sees will be its own mother. But if, by chance, a porcupine wanders through the gosling's visual field during the imprinting period, the gosling will imprint on the porcupine. In this case, we insist, the imprinting mechanism is working exactly as designed, since it's designed to *imprint on the first large moving object you see*. True, something isn't going according to evolution's plan, but we shouldn't use the term "dysfunction" to describe that sort of wrong, as it would confuse many issues that deserve to be kept separate.

In sum, I'm happy to acknowledge that the move from seeing delusions as psychologically beneficial in some way, to the conclusion that they are evolutionary adaptations—either products of biological design, or outputs of systems working just as designed—is not beyond dispute. But I think it would be quite premature to take the adaptationist view off the table.

3. Response to Núñez de Prado Gordillo

3.1 A scaffolding for Mad Pride

Núñez de Prado Gordillo centers his commentary on the political dimension of my book, particularly on the connections between madness-as-strategy and Mad Pride. I am grateful for the opportunity to expand my view slightly, as I only touch upon it in passing in the book. As I note there, there are four ways in which the dysfunction/strategy distinction can be useful: in its implications for history, philosophy, treatment, and Mad Pride:

> [T]his attempt to retrieve madness-as-strategy as a coherent way of seeing contributes to the project of providing intellectual scaffolding for the emerging movement variously known as Mad Pride, mad resistance, or mad activism. (Garson 2022a, 12)

I also write that, in the view of madness I endorse, "madness is not always a disease to be cured but a force of disruption to be reckoned with".

I agree that these remarks are cryptic at best, and I sought to elaborate on them in another paper, "Madness and Idiocy" (Garson 2023). I'll recapitulate the core idea here. My view is that, to the extent that Mad Pride takes inspiration from movements like Gay Pride, Black Pride, or Deaf Pride, it must start by rejecting a dominant cultural narrative about what it means to be gay, Black, or deaf. The very idea of Gay Pride makes little sense if one really thinks that being gay amounts to having a pathology of sexual orientation. Similarly, Deaf Pride makes little sense if one thinks that being deaf is merely a disease that the world would be better off without. The most obvious way to dislodge such pathologizing paradigms would be to replace them with an alternative paradigm that frames the issue in a more positive, empowering way. For example, one might say that being gay is just a variation in sexual orientation, like having freckles or a chin cleft, and variation is something to be celebrated rather than eliminated. Similarly, one might emphasize the way in which being deaf involves not (merely) a dysfunction, but an entirely different mode of engaging with the world that has value in its own right.

By the same token, if we want to promote Mad Pride, we must begin with rejecting the dominant cultural narrative that holds that the mad mind is a broken mind—a mind that fails to work as it should. That medical framing invites us to see madness—or its medically sanitized cousin, "mental disorder"—as a disease to be treated. At most, that framing gives us "mental health advocacy", but falls short of Mad Pride. One way to shift the conversation away from pathologizing framings is simply to offer an alternative framing, one that's more positive and empowering. Madness-as-strategy provides one such framing. As I put it, once we have the concept of madness-as-strategy at our disposal, "madness-as-dysfunction can no longer be a silent default in approaching the mad" (Garson 2022a, 260). For this reason, I called madness-as-strategy part of the "intellectual scaffolding" of Mad Pride.

I still agree with my assessment. But in retrospect, I think "scaffolding" is not an entirely apt metaphor, for two reasons. First, "scaffolding" is a temporary affair. It's not a permanent part of the foundation of the building, but something to be thrown out after it has served its function. But I think of madness-as-strategy as a resource that we can draw on permanently, not merely as a conceptual tool to get us from one place to another, intellectually speaking. (It is not in that respect like Wittgenstein's ladder, to be "thrown away" after it does its job.) In another way, however, "scaffolding" is too ambitious. Scaffolding is *indispensable* for constructing a building, like flour is indispensable for baking a cake. But I'm open to the possibility that there may be other empowering alternatives to the standard pathologizing framework. Perhaps a better way to describe madness-as-strategy is as a support beam for Mad Pride. It is part of the foundation of a building, but not the only thing that keeps the building up. Of course, to depict madness-as-strategy as a support beam for Mad Pride raises the question of what other cogent alternatives there are for thinking about madness in a more empowering and positive way. What are some other "support beams"? Núñez de Prado Gordillo suggests that the emerging neurodiversity paradigm is a valuable alternative support beam. (He also sketches an alternative view, madness-as-right. Intriguing as it is, I will not have space to explore it here.) To suggest that madness-as-strategy and the neurodiversity paradigm are different, non-pathologizing alternatives to the default medical model, however, invites us to look more closely at how they relate to one another.

3.2 Madness-as-strategy and the neurodiversity paradigm

The rest of Núñez de Prado Gordillo's commentary explores points of convergence and divergence between madness-as-strategy and the neurodiversity paradigm. I think he highlights real points of intersection and tensions between the two views. But before diving into these, I feel the need to clarify, a bit more rigorously, what I mean by madness-as-strategy and to highlight the diversity of specific approaches that fall under that general framework, as I believe he may be construing it too narrowly. This narrow construal is partly responsible for the appearance of tension.

Núñez de Prado Gordillo suggests that madness-as-strategy is equivalent to the idea that madness is a "natural reaction" to adversity, as memorably described by Shaughnessy, or as a "sane response to an insane society", as Laing put it. He's correct that those are the primary contemporary examples I focus on in my book. But the idea of madness as a "natural reaction" is merely one expression of madness-as-strategy among others. (For example, for Robert Burton, melancholy is a strategy, but it is not clearly a "natural response" to the trials of life, but a divine wake-up call.) In other places, I elaborate this point in more detail (Garson 2024b), where I envision three main expressions of this view in contemporary mental health:

1. **Madness as a strategy for coping with present-day adversity.** This expression comes closest to seeing madness as a "natural reaction". For instance, delusions of grandeur may be a way of coping with (and hence a "natural reaction to") a sense of insignificance in life (Isham et al. 2022). Depression may represent the mind's designed signal that the organism is in an untenable situation—such as a problematic relationship, career, or social setting (Nesse 2019). Khalidi's commentary centers on Fanon, who, as a psychiatrist in French-occupied Algeria, often characterized the mental health struggles of his patients, like paranoia and aggressiveness, as not just natural (in the sense of expectable) responses to colonialism, but in some sense functional responses.
2. **Madness as a strategy for coping with past adversity.** For example, one way of conceptualizing borderline personality disorder (BPD) is as a set of survival strategies developed in response to trauma, rather than as evidence of a brain defect (Brüne 2016). Similarly, for some individuals, the experience of voice-hearing can represent dissociated contents striving for reintegration (Longden and Read 2017). In this light, movements like the Power Threat Meaning Framework invite us to ask not "what's wrong with you", but "what happened to you?" (Johnstone 2022).
3. **Madness as an evolved cognitive strategy for benefiting the group.** This includes cognitive traits typically associated with neurodiversity, such as ADHD, autism, and dyslexia, which may have evolved due to their group-level benefits. For instance, there is evidence suggesting that dyslexia and ADHD represent evolved cognitive styles that offer unique advantages to communities (Taylor and Vestergaard 2022; Hunt and Jaeggi 2022; Garson 2022c). This is still a strategy, but not at the level of the individual, but the group: it's a group-level strategy for community survival.

The last of these three dovetails neatly, I think, with what Núñez de Prado Gordillo calls the "relational-ecological" model of neurodiversity, in which conditions like ADHD are construed as forms of cognitive diversity that "might be an adaptive feature for maximizing *collective* thriving and fitness". So construed, there's no discrepancy between madness-as-strategy and neurodiversity. This relational-ecological approach to neurodiversity can be seen as one expression of madness-as-strategy.

3.3 Does madness-as-strategy support the "normalcy" paradigm?

So far, Núñez de Prado Gordillo and I seem to agree. But he rightly identifies tensions between madness-as-strategy and certain formulations of neurodiversity. In particular, he raises the question of whether madness-as-strategy subtly reinforces the normalcy paradigm that the neurodiversity paradigm seeks to dismantle. As he puts it, madness-as-strategy seems to assume that there is

> [S]ome essential assortment of mental functions and capacities that conform to a natural or universal standard of *normal* cognitive functioning; a fixed mold into which madness must fit if we are to see purpose, value, and an enactment of human cognitive potential in it. (Núñez de Prado Gordillo this issue, 50)

This is a complex question, and I don't hope to entangle all its threads here. My current view is that there is a thick sense of normalcy that I reject, and a thin sense of normalcy that I accept. Put differently, I agree that my view could probably ground a rather thin sense of normalcy—but I do not think that is a bad thing.

Let me try to elaborate on what I take to be the core objection. We have a culturally-conditioned idea that certain ways of being or acting are *normal* (natural, good), while others are *abnormal* (unnatural, bad). For example, we tend to think that experiencing a moderate degree of low mood after losing a job is normal, but that experiencing incapacitating depression as a result of the same loss is abnormal. Moreover, the idea goes, medicine has a special responsibility to help people achieve normalcy.

Given all this, one might argue that the idea of normalcy is harmful, particularly for those who, by current social standards, are deemed "abnormal". We might be better off without this idea and ought to be wary of philosophical attempts to reinstate it. Amundson (2000) articulates this line of thought quite elegantly.

With these ideas in mind, one might worry that madness-as-strategy is not a way of challenging normalcy but of reinstating it, and perhaps even giving it a solid philosophical foundation. To get there, I'll use the example of depression and give a bit of background. I have argued that we should begin to see depression not as a pathology, but as a functional, adaptive, and perhaps even evolved response to adversity—something like the brain's designed signal that something in the environment isn't going well and needs more attention. I believe this paradigm shift is incredibly beneficial for both treatment and stigma (Garson 2024b). First, it helps us reorient treatment. If depression is a designed signal that something in one's life isn't going well, then it stands to reason that we should listen to what it's trying to say, rather than bombard it with antidepressant drugs. Second, there's emerging evidence that simply framing depression as a designed signal (rather than a chemical imbalance) benefits patients, as it gives them a greater sense of optimism about recovery (Kneeland, Schroder, and Garson in prep). Given these benefits, I actually think it would be morally pernicious for mental health providers *not* to make patients aware of this perspective (Garson forthcoming).

Of course, if there are functional forms of depression, then there may be dysfunctional forms as well—and indeed, they do seem to exist. Depression can arise not as a designed signal that something is wrong in one's life, but as a consequence of a brain tumor or neurodegenerative disease. In these cases, depression may still have some value—for instance, as a diagnostic indicator or as an opportunity for personal reflection—but it doesn't have the same *sort* of value that the functional kind has. (Recall that I do not entirely reject madness-as-dysfunction, and as I state in my book, I think there are likely cases in which this concept deserves to be applied—see Garson 2022a, 2).

Now, back to normalcy: I wouldn't object if someone wanted to use the term "normal" for what I describe as functional. Millikan (1984) sometimes uses the term in this way. For her, the "normal" (or "proper") function of, say, the kidney is to eliminate waste. Along those lines, I wouldn't object if someone wished to describe depression as a "normal response" to certain kinds of adversity, such as social humiliation, if what they mean is that it is functional, i.e., it represents everything in the cognitive system working as designed. Using "normal" in this sense, we might also say that ADHD is an example of normal cognitive variation—it represents one way that the mind is designed to work, rather than a deviation from design. In contrast, a neurodegenerative disease like Alzheimer's isn't an example of "normal cognitive variation"; instead, it represents dysfunctional or pathological variation. Not only do I find this notion of normalcy unproblematic, but it's hard for me to even imagine what biomedicine or psychiatry would look like if it tried to do without it.

Of course, sometimes the notion of "normalcy" is used in a much thicker, metaphysically and ethically inflated sense—to denote a property that is meant to be at once universal, natural, and inherently good. For example, if someone says that same-sex sexual attraction is "unnatural", they presumably mean it in this thicker sense. I don't believe in normalcy in that sense. Tumors, for example, are perfectly "natural", in that they arise through the same natural processes as any other biological development. Moreover, there is nothing inherently good about a trait performing its function. Teen pregnancy, for instance, represents normal reproductive function, but it's still something society might wish to discourage.

In short, I believe that there is a thin notion of normalcy, which is pretty much synonymous with proper function—a concept that is relatively unproblematic and may even be central to the aims of biomedicine and psychiatry. In contrast, the thicker, metaphysically and ethically inflated notion of normalcy does not correspond to anything real, and society would likely be better off without it. But it would not be fair to try to saddle madness-as-strategy with this thicker notion of normalcy.

4. Response to Jeppsson and Lodge

4.1 The allure of madness

I'm incredibly grateful to Jeppsson and Lodge for their unique contributions to this symposium. Starting with the idea that madness can be a strategy—the mind's wake-up call, a coping mechanism, and so on—they raise an important question: why couldn't a mad person consciously choose to implement various strategies for dealing with their madness? In other words, once we consider that madness may be a strategy, it makes perfect sense to begin discussing the strategies we can consciously adopt to navigate mad experiences.

The theme that struck me most from their piece, particularly in their personal testimonies of madness, is the theme of madness's *allure*—a theme I will briefly develop before addressing their individual narratives. Madness, when presented to us in the medically sanitized and safe guise of "mental disorder", is almost universally framed in biomedical literature as something that *happens* to us. As I put it in the book, it is "an accident that happens from time to time and that tragically befalls an otherwise healthy person, a promising young man or woman" (Garson 2022a, 263).

Arguably, this idea of passivity in relation to madness has been a cornerstone of the biomedical movement since its emergence in the 1980s, which became solidified in the 1990s, the so-called “Decade of the Brain”. One popular slogan at the time held that “depression is just like diabetes” or “schizophrenia is just like cancer”. Nobody chooses cancer or diabetes; they happen to us. This idea that we are passive in relation to madness remains central to many advocacy movements. For example, the National Alliance on Mental Illness (NAMI), in their anti-stigma tips, tell us that we should “encourage equality between physical and mental illness”.[2] And as a recent billboard campaign sponsored by a mental health advocacy group, Bring Change 2 Mind, reminds us: “Imagine if you got blamed for having cancer”.

Within this political context, discussing the allure of madness is provocative, to say the very least. Some might even call it dangerous. After all, describing madness as alluring suggests we have a choice in relation to it. To depict madness as a temptation insinuates that we might have some capacity to choose—that we may have a certain degree of freedom with respect to our madness. Will we resist the temptation of madness or succumb to it? As Jeppsson and Lodge emphasize, many mad people confront this dilemma in a concrete way—for example, when considering whether to continue taking antipsychotic or mood-stabilizing medications, or to stop. This raises the prospect that the mad person could, in some sense, be morally accountable for their madness. Hence the risk.

Along with the risk, however, there is tremendous opportunity: by depicting madness as a temptation and foregrounding the role of choice, we highlight the role of agency, which some might read as empowering. The mad person is no longer the passive subject of their madness but rather has an opportunity to exercise agency in relation to it, and to select new strategies for navigating it.

I don’t have particularly insightful suggestions for steering through this fraught set of concepts: agency, blame, madness. However, I do believe that this is among the most urgent tasks confronting philosophers of psychiatry and madness, and I applaud Jeppsson and Lodge for bringing this to our attention.

I want to spend time reflecting on Lodge’s and Jeppsson’s accounts of their own madness, which I found rich and philosophically fruitful. This is not a matter of critique but of extending the conversation. My own view, which aligns with theirs, is that the exclusion of mad narratives from the

[2] https://www.nami.org/education/9-ways-to-fight-mental-health-stigma/

discipline of philosophy—a discipline that always seems to exist at the precipice of madness—has deeply impoverished our field (Kusters 2020 develops this theme at length). I hope these brief comments will serve as a continued stimulus for new growth.

4.2 Manic subjectivity and hermeneutic injustice

Lodge uses his own experience of mania to home in on two aspects of mania that clinical classifications typically exclude. The first is what he describes as the expansion of manic subjectivity. This, he notes, can be read as a philosophical redescription of what's colloquially described as an "inflated sense of self"—a sense of being confronted with more stimuli and ideas than one could possibly attend to. The expansion of manic subjectivity leads to the second aspect of mania: the individual seeks to render this enriched state of consciousness intelligible by drawing upon their existing, albeit crude, conceptual toolkit. Because our culture does not provide us with a sufficiently expansive toolkit for making sense these experiences, people often turn to conceptualizations deemed strange, bizarre, mystical, or even "delusional".

For example, one rather obvious concept within our culturally-conditioned toolkit is messianism. Perhaps "manic subjectivity" describes what Jesus felt in the aftermath of his baptism by John, or during the Transfiguration, when Moses and Elijah themselves appeared in their glory to provide a cosmic download of information. (Richard Saville-Smith (2023) develops these religious themes brilliantly.) The notion that the interpretation of one's manic subjectivity is somehow limited by one's impoverished conceptual toolkit also echoes Sullivan-Bissett's (2018) discussion of "one-factor" theories of delusions (much as I hesitate to use the term "delusion" to describe such exalted states of mind).

This idea that manic subjectivity must be articulated, however imperfectly, through a limited conceptual toolkit also offers a novel way of considering epistemic injustice—particularly the form of epistemic injustice that Fricker (2007) calls "hermeneutic injustice", in contrast to testimonial injustice. Whereas testimonial injustice occurs when someone's status as a knower is denigrated, hermeneutic injustice arises when someone is unfairly deprived of the concepts necessary to articulate their experiences. (Fricker's primary example is the absence of the concept of sexual harassment in the 1950s and 1960s, which left victims without a way to accurately identify the specific wrong inflicted upon them.)

Viewed in this light, the connection between madness-as-dysfunction and hermeneutic injustice becomes apparent. By offering only a single dominant narrative to make sense of manic experiences—dysfunction, or "something seriously wrong in the mind"—we unfairly deprive people of alternative conceptualizations that might better serve their needs, both psychologically and existentially. (I explore these in Garson 2025c). These alternatives include trauma-centered explanations, internal family systems (IFS) models, spiritual paradigms, and others. Not only does the predominance of dysfunction-centered framings in psychiatry deprive us of these meaning-making alternatives, but it actively discourages them.

Finally, I want to highlight a connection that I'm sure Lodge has already considered, between the kind of crisis induced by a manic experience and the sense of forlornness that follows an LSD trip or other powerful psychedelic experience. How does one come to terms with that experience? How does one construct a metaphysical and ethical worldview adequate to it? Similar conversations took place in the United States during the height of LSD culture in the early 1960s (e.g., Lee and Shlain 1992; also see Kusters 2020).

4.3 Back to madness's allure

Jeppsson has used her experiences to reflect on basic questions of epistemology, particularly the problem of external world skepticism (2022a; 2022b). How can we know that the external world—the mainstream world with its trees, houses, cows, grass, and the like—is real? Philosophers have invented a whole range of strategies to reassure themselves that the external world is, in fact, real, solid, and substantial, just as it appears to be.

I have learned quite a lot from Jeppsson's work about the various strategies—no pun intended—that philosophers have employed to achieve such reassurance. One such approach, the Wittgensteinian strategy, posits that belief in the external world is not something that can be logically demonstrated but rather serves as a foundation for reasoning itself. Some epistemologists describe belief in the external world as a "hinge belief", a special kind of belief that makes reasoning about anything possible (Pritchard 2021). To put this hinge belief into question is either incoherent (as Wittgenstein 1969 suggests in *On Certainty*) or self-defeating. Jeppsson has performed an invaluable service for philosophy by illustrating, through her experience of the demon world, that questioning the existence of the external world is neither incoherent nor self-defeating for reason. Moreover, I believe she has successfully argued that most attempts to respond to external world skepticism are, in one way or another, question-begging, in the sense that they rely on assumptions that the skeptic couldn't reasonably accept.

What was particularly striking to me about Jeppsson's testimony here, however, is the way she ties her reflections to the problem of madness's allure. As she notes, unlike the delusional belief that a famous actress is in love with me, there doesn't seem to be anything particularly alluring about the prospect that I'm living in a demon world. I think most of us, if presented with that possibility in the abstract ("If you take the red pill, you'll live under the conviction that murderous demons are persecuting you") would choose not to.

Yet, behind this initial revulsion lies a deeper allure. Who doesn't want to be the main character in a cosmic drama? To occupy the center of a persecution narrative is to be *somebody*, and for many, being somebody is better than being nobody. Jeppsson's insight, I suspect, could prove fruitful in helping us understand why so many people, when confronted with the following two possibilities—(1) nobody is after you; you simply have a brain disorder that makes you think people are after you versus (2) everybody is actually after you, and the sooner you embrace this truth, the more likely you are to survive—choose the second option. I am thinking here primarily about targeted individuals, that is, people who believe themselves to be victims of gang stalking or electronic harassment (e.g., Garson 2024a, Garson 2024c). In 2016, *The New York Times* estimated that there are about 10,000 people who identify as targeted individuals, but the number now is probably much higher. I believe we cannot entirely understand this puzzling sociological phenomenon without drawing upon Jeppsson's and Lodge's notion of allure.

Ultimately, what I want to emphasize about Jeppsson's and Lodge's insightful paper is that they truly demonstrate the payoff of mad philosophy. They illustrate how madness can serve as a disruptive force for philosophy. This is simply a more roundabout way of affirming that madness, indeed, has its benefits.

REFERENCES

Abed, Riadh, and Paul St. John-Smith, eds. 2022. *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge: Cambridge University Press.

Amundson, Ron. 2000. "Against Normal Function." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 31 (1): 33–45.

Bortolotti, Lisa. 2023. *Why Delusions Matter*. London: Bloomsbury.

Brüne, Martin. 2016. "Borderline Personality Disorder: Why 'Fast and Furious'?" *Evolution, Medicine, and Public Health* 1: 52–66.

Engel, George L. 1977. "The Need for a New Medical Model: A Challenge for Biomedicine." *Science* 196: 129–136.

Fagerberg, Harriet, and Justin Garson. 2024. "Proper Functions Are Proximal Functions." *British Journal for the Philosophy of Science*. https://doi.org/10.1086/731869

Fineberg, Sarah K., and Philip R. Corlett. 2016. "The Doxastic Shear Pin: Delusions as Errors of Learning and Memory." *Cognitive Neuropsychiatry*. 21 (1): 73–89.

Foucault, Michel. 1977. "Nietzsche, Genealogy, History." In *The Foucault Reader*, edited by Paul Rabinow, 76-100. New York: Pantheon Books.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Garson, Justin. 2022a. *Madness: A Philosophical Exploration.* New York: Oxford.

Garson, Justin. 2022b. *The Biological Mind, Second Edition.* London: Routledge.

Garson, Justin. 2022c. "Dyslexia: Beyond a Disorder." *Psychology Today,* November/December, 26–27. https://www.psychologytoday.com/gb/blog/the-biology-of-human-nature/202207/seeing-dyslexia-as-a-unique-cognitive-strength-rather-than

Garson, Justin. 2023. "Madness and Idiocy: Rethinking a Basic Problem of Philosophy of Psychiatry." *Philosophy, Psychiatry, Psychology* 30(4): 285–295.

Garson, Justin. 2024a. "From Pathology to Purpose: Targeted Individuals and the Harms of Psychiatry's Dysfunction Paradigm." In *Gangstalking: Academic Intersections and Ethical Issues,* edited by L. B. Johnstone, 76–92. Cambridge: Ethics International Press.

Garson, Justin. 2024b. "Madness-as-Strategy as an Alternative to Psychiatry's Dysfunction-Centered Model." In *Theoretical Alternatives to the Psychiatric Model of Mental Disorder Labeling: Contemporary Frameworks, Taxonomies, and Models,* edited by A. Cantú, E. Maisel, and C. Ruby, 406–423. Cambridge: Ethics International Press.

Garson, Justin. 2024c. "Targeted." *Aeon,* September 2, 2024. https://aeon.co/essays/how-the-psychiatric-narrative-hinders-those-who-hear-voices

Garson, Justin. Forthcoming. "Beyond Psychiatry: Rethinking Madness Outside Medicine." In *Madness and Mental Health,* edited by Edward Harcourt. Cambridge: Cambridge University Press.

Gluckman, Peter, Alan Beedle, and Mark Hanson. 2009. *Principles of Evolutionary Medicine.* Cambridge: Cambridge University Press.

Hunt, Adam D., and Adrian V. Jaeggi. 2022. “Specialised Minds: Extending Adaptive Explanations of Personality to the Evolution of Psychopathology.” *Evolutionary Human Sciences* 4: E26.

Isham, Louise, et al. 2022. “The Meaning in Grandiose Delusions: Measure Development and Cohort Studies in Clinical Psychosis and Non-Clinical General Population Groups in the UK and Ireland.” *The Lancet Psychiatry* 9 (10): 792–803.

Jeppsson, Sofia. 2022a. “My Strategies for Dealing with Radical Psychotic Doubt.” *Schizophrenia Bulletin* 49(5): 1097–1098.

Jeppsson, Sofia. 2022b. “Radical Psychotic Doubt and Epistemology.” *Philosophical Psychology* 36 (8): 1482–1506.

Jester, Dylan J., et al. 2023. “Review of Major Social Determinants of Health in Schizophrenia-Spectrum Psychotic Disorders: I. Clinical Outcomes.” *Schizophrenia Bulletin* 49 (4): 837–850.

Johnstone, Lucy. 2022. *A Straight Talking Introduction to Psychiatric Diagnosis.* PCCS Books.

Kneeland, E., Schroder, H, and Garson, Justin. In Prep. “An Ambiguity in the Notion of a “Biomedical” Approach to Mental Health.”

Kusters, Wouter. 2020. *A Philosophy of Madness: The Experience of Psychotic Thinking.* Cambridge, Mass: MIT Press.

Lancellotta, Eugenia. 2022. “Is the Biological Adaptiveness of Delusions Doomed?” *Review of Philosophy and Psychology* 13: 47–63.

Lee, Martin A., and B. Shlain. 1992. *Acid Dreams: The Complete Social History of LSD: The CIA, the Sixties, and Beyond.*

Longden, Eleanor, and John Read. 2017. “‘People with Problems, Not Patients with Illnesses’: Using Psychosocial Frameworks to Reduce the Stigma of Psychosis.” *The Israel Journal of Psychiatry and Related Sciences* 54(1): 24–28.

López-Silva, Pablo. 2023. “Minimal Biological Adaptiveness and the Phenomenology of Delusions in Schizophrenia.” In *The Philosophy and Psychology of Delusions: Historical and Contemporary Perspectives*, edited by Ana Falcato and Jorge Gonçalves, 126–140. New York: Routledge.

Millikan, Ruth. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Nesse, Randolph M. 2019. *Good Reasons for Bad Feelings: Insights from the Frontier of Evolutionary Psychiatry.* Dutton.

Pritchard, Duncan. 2021. “Wittgensteinian Hinge Epistemology and Deep Disagreement.” *Topoi* 40 (5): 1117–1125.

Queloz, Matthieu. 2021. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering.* Oxford: Oxford University Press.

Read, John. 2005. “The Bio-Bio-Bio Model of Madness.” *The Psychologist* 18 (10): 596–597.

Ritunnano, Rosa, et al. 2022. “Subjective Experience and Meaning of Delusions in Psychosis: A Systematic Review and Qualitative Evidence Synthesis.” *The Lancet Psychiatry* 9 (6): 458–476.

Saville-Smith, Richard. 2023. *Acute Religious Experiences: Madness, Psychosis, and Religious Studies.* London: Bloomsbury.

Sober, Elliot., and David S. Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Cambridge, MA: Harvard University Press.

Soper, C. A. 2018. *The Evolution of Suicide.* Dordrecht: Springer.

Sullivan-Bissett, Ema. 2018. “Monothematic Delusion: A Case of Innocence from Experience.” *Philosophical Psychology* 31 (6): 920–947.

Swanepoel, Annie, and C. A. Soper. 2025. “Mental Disorders May Prevent, Not Cause, Suicide.” *BJPsych Bulletin* 49 (2): 123–125.

Szasz, Thomas. 1960. “The Myth of Mental Illness.” *American Psychology* 15: 113–118.

Taylor, Helen, and Martin D. Vestergaard. 2022. “Developmental Dyslexia: Disorder or Specialization in Exploration?” *Frontiers in Psychology* 13: art. 889245.

Trivers, Robert. 2011. *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life.* New York: Basic Books.

Wakefield, Jerome C. 1999. “Mental Disorder as a Black Box Essentialist Concept.” *Journal of Abnormal Psychology* 108: 465–472.

Wittgenstein, Ludwig. 1969. *On Certainty.* Oxford: Blackwell.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 2 | 205-228
https://doi.org/10.31820/ejap.21.2.6

HABITS AND DISPOSITIONS IN FRANK RAMSEY'S PHILOSOPHY

Alice Morelli[1]

[1] Ca' Foscari University, Venice, Italy

Original scientific paper – Received: 12/02/2025 Accepted: 27/07/2025

ABSTRACT

This paper examines Ramsey's use of the concepts of habit and disposition, challenging the common interpretation that he employs them interchangeably in his theory of belief. This interpretative trend reflects a broader tendency to equate habit and disposition, based on the assumption that a habit is an acquired disposition to act. However, the precise relationship between these concepts often remains underexplored and it is not clear whether habits are merely a subset of dispositions or if they are conceptually distinct. Using Ramsey's writings as a case study, this paper argues that their relationship is more nuanced than a reductive equivalence suggests. I advance a twofold thesis: first, I argue that Ramsey's use of the notions of habit and disposition is more complex than typically assumed, as he employs them in distinct philosophical contexts and conceptualizes them in different ways. Second, I distinguish between a logical-grammatical kind of dispositionalism and a metaphysical one to argue that the notion of habit is dispositional but habits are not metaphysically equivalent to dispositions. Ramsey conceptualizes habits as methods, rules, procedures of thought, whereas dispositions are understood as tendencies or inclinations engendered and shaped by habits.

Keywords: Frank Ramsey; habit; disposition; pragmatism; normativity.

Correspondence: alice.morelli@unive.it

1. Introduction

Is stating that beliefs are habits of mind the same as claiming that beliefs are dispositions to act? In other words, are the concepts of habit and disposition equivalent? I will argue against this equivalence by looking at Frank Ramsey's theories of belief and judgment.

Ramsey's emphasis on the notion of habit represents a distinctively pragmatist strand within his philosophy (Misak 2016; Tuzet 2020), one that he explicitly develops by drawing on Bertrand Russell's *The Analysis of Mind* (1921/2008) and Charles Peirce's writings.[1] The notion of habit plays a central role in Ramsey's causal theory of belief, which the literature commonly characterizes as "dispositional". According to this account, a belief is defined by its role in guiding behavior: it is a mental state—a disposition—that produces actions in conjunction with desires. Although closely related, Ramsey commentators frequently treat the terms "habit" and "disposition" as interchangeable, using them synonymously to describe the principles underlying action. On this view, beliefs are understood by Ramsey as habits *or* dispositions to act (Misak 2016, 2022; Engel 2005; Kraemer 1985). As Soroush Marouzi notes, "Ramsey refers to these dispositions as habits", whose nature aligns with the moderate behaviorism characteristic of several philosophers of the 1920s, including Ralph Perry, Edwin Holt, and Edward Tolman (see Marouzi 2024, 9).[2]

This interpretative trend reflects a broader inclination to conflate habit and disposition, grounded in the view that a habit is an acquired disposition to act. Already in Aristotle, habit (*hexis*) is defined as an acquired disposition that enhances the agent's performance.[3] Subsequent thinkers in the Aristotelian tradition have continued to conceptualize habit within the framework of disposition. Thomas Aquinas, for instance, describes habit as an acquired disposition that is resistant to change and rooted in stable causes (Miner 2013). Similarly, Félix Ravaisson (2008) defines habit as a disposition relative to change.

In contemporary philosophy, the relation between habits and dispositions is frequently invoked to support the idea that, once acquired, a habit disposes an agent to act in particular ways. This perspective serves as a conceptual counterpoint to behaviorist accounts that characterize habits as

[1] Ramsey read *Chance, Love and Logic*, a volume published by Odgen in 1923 as part of his *International Library* series. It reprinted six articles that Peirce had published in the *Popular Science Monthly* between 1877 and 1878. For a detailed reconstruction of Ramsey's pragmatist sources, see Misak (2016).

[2] This kind of behaviorism is discussed in Mills (1998).

[3] For a detailed account of the Aristotelian theory of habit, see Lockwood (2013) and Chiaradonna and Farina (2022).

automatic, conditioned responses to stimuli. In contrast to these reductive views, authors such as Daniel Hutto and Ian Robertson (2021) argue that habitual doings display intelligence, understood as focused and flexible, world-directed dispositions. Along similar lines, Katsunori Miyahara, Taller Ransom, and Shaun Gallagher (2021) introduce the notion of "skilful dispositions" to describe the enduring tendencies that underlie habitual yet attentive forms of action. Contemporary efforts to rehabilitate the concept of habit essentially draw on pragmatist thought, which has long advocated for a more nuanced understanding of the term. From Wiliam James's depiction of individuals as "bundles of habits" (James 1914) to John Dewey's claim that "habit means will" (Dewey 2023, 21), pragmatism portrays habits as dynamic, ecological, and self-organizing structures. In this framework, rather than opposing reflective awareness, habits mediate between pre-reflective and reflective processes, revealing their integral role in intelligent, situated action.[4]

Despite their frequent use, the concepts of habit and disposition are often employed interchangeably in the literature, with little clarification of their conceptual relationship. If habits are defined as dispositions to act, does this mean that they are merely a subset of dispositions, or do they possess distinct conceptual features? While I am not suggesting that the aforementioned authors endorse a form of metaphysical reductionism, I do believe that this lack of terminological precision risks obscuring important philosophical distinctions. This paper aims to address this gap by providing a careful examination of the distinction between habits and dispositions in Ramsey's philosophy. I take this to be the paper's most significant contribution to the current scholarly discourse. The question, as I frame it, is not whether habits are dispositions or vice versa, but how the conceptual relationship between the two is best understood.

I argue that Ramsey's treatment of the concepts of habit and disposition is more nuanced than is commonly assumed. Although closely related, these notions are employed in distinct theoretical contexts and serve different philosophical purposes in his work. To support this claim, I begin by demonstrating that the concepts of habit and disposition are not reducible to one another, as evidenced by the different ways Ramsey deploys and conceptualizes them across various philosophical contexts. I then examine the claim that habits are dispositions by distinguishing between two forms of dispositionalism: a logical-grammatical approach and a metaphysical one. I argue that mental habits are dispositional insofar as they give rise to

[4] Other notable exceptions to the narrow view of habit are found in sociology and in the phenomenological tradition (Merleau-Ponty 2012). For instance, Pierre Bourdieu employs the notion of *Habitus* precisely "to set aside the common conception of habit as a mechanical assembly or preformed programme" (Bourdieu 1977, 218).

beliefs and opinions, which belong to the logical category of the dispositional. These beliefs, once formed through habitual thought, dispose the agent toward particular patterns of behavior. I conclude that although the notion of habit is dispositional, habits are not metaphysically identical with dispositions. Rather, mental habits—or habits of thought—should be understood as *methods* of thought, whereas dispositions are *tendencie*s or *inclination*s shaped and structured by these habits. Furthermore, I suggest that this conceptual clarification sheds light on the distinction that Ramsey articulates in *General Propositions and Causality* [**GPC**] between judgments and rules for judging (Ramsey 1994, 149)—a distinction that remains a point of interpretative contention in the literature.[5]

2. Against conceptual reduction: Context*s*

In this section, I argue that although Ramsey acknowledges important similarities between the concepts of habit and disposition, he also preserves key distinctions that preclude their conceptual reduction.[6] Specifically, he employs these notions in distinct philosophical contexts and for different theoretical purposes. Conceptual reduction—as I use the term here—occurs when one concept (A) is defined entirely in terms of another (B). I will contend that Ramsey neither reduces habit to disposition nor disposition to habit.

Despite their differences, habits and dispositions share three salient features. First, they both *govern* human actions and behavior, serving as *explanatory* principles. In this sense, they are both principles of action. In *Truth and Probability* [**TP**], Ramsey characterizes habit as a rule or law of behavior—one of the general principles according to which the human mind functions (Ramsey 1994, 90). Similarly, in *On Truth* [**OT**], he characterizes the dispositional as the persistent background of the mind that is manifested in action and thought, and invoked to explain specific instances of each (Ramsey 1991, 43). Second, both are manifested in *particular* acts. While a habit is a general rule of action, it leads to specific thoughts and behaviors. Likewise, a disposition reveals itself through its manifestations—or through the actions that *would* occur under suitable conditions. Ramsey notes that dispositions are not themselves acts of thought, but are manifested in such acts (ibid., 98-99). Third, both contribute to the *explanation* of immediate or spontaneous action—that is,

[5] See Holton and Price (2003), Misak (2016), Marouzi (2024), and Marion (2012).

[6] There can be at least three kinds of reduction (McKitrick 2009, 32): conceptual, epistemic, and metaphysical. In this paper, I argue that Ramsey does not commit to conceptual reduction. However, I will suggest that part of his general framework aligns with the idea that habits and dispositions are not reducible to one another from a metaphysical perspective too.

action performed without conscious deliberation. In this respect, dispositions can contribute to the explanation of habitual behavior. Ramsey illustrates this point in **OT** using the example of believing that the Cambridge Union is located on Bridge Street. This belief is dispositional in that it exists as a potentiality: it may guide action even when not actively entertained. Though rarely formulated as an explicit thought, the belief often manifests in action—for example, "by my turning my steps that way when I want a book from the Union Library" in Cambridge ("where I am at home, I go there habitually without having to think": ibid., 44–45). This action, guided by a dispositional belief, is executed "without any process of thought", whereas in other contexts the philosopher might need to reflect on the Union's location. Most importantly, Ramsey does not infer that habitual action is unintelligent. On the contrary, he contends that such actions bypass specific acts of thought because habit "makes the intermediate state of thought disappear" (ibid., 51). Habit formation, he writes, "telescopes" thought: it eliminates the intermediary stage of judgment, enabling conditions to give rise directly to action.

Despite these similarities, the notions of habit and disposition differ in two significant ways: they are used in distinct philosophical contexts and are conceptualized differently.[7] This section addresses the first difference; the second will be explored in the next section.

Ramsey assigns a central role to habit in his causal theory of belief, while disposition is pivotal in his classification of mental states. When discussing habits, he focuses on measuring degrees of belief; when analyzing dispositions, he aims to clarify the nature of judgment. It is important to note that in *Facts and Propositions* [**FP**], belief and judgment are treated as synonymous (Ramsey 1994, 34), but this identification is explicitly rejected in **OT**.[8] There, beliefs are categorized as dispositional states, while judgments are construed as spatio-temporal, affirmative acts of thought with propositional content. Dispositions and acts thus fall under distinct logical types. Consequently, Ramsey does not develop a dispositional theory of judgment, though he arguably advances a dispositional theory of belief.

[7] More precisely, Ramsey uses the two notions as synonyms on one occasion in **OT** while clarifying the nature of judgment. He writes: "(…) clearly the same mental disposition or habit was manifested the day before he slapped what was really Jones' back" (Ramsey 1991, 49). However, this singular occurrence does not conflict with the idea that there are more substantial differences between the two notions.

[8] They are equivalent in that they cover the range of mental states from conjecture to knowledge. In **OT**, Ramsey (1991, 8) uses the terms "belief" and "judgment" interchangeably when analyzing the ascriptions of "true" and "false" to mental states, but he specifies that their ordinary meaning is narrower.

Cheryl Misak (2016) claims that Ramsey's dispositional theory of belief as a habit is already present in **FP**. While I agree that a dispositional account is articulated there in substance, it is noteworthy that Ramsey uses neither the term "habit" nor "disposition" in that text. These terms appear more explicitly in **TP**, *Knowledge* [**K**] and **GPC**. They have only a limited presence in **OT**, where "habit" appears sparingly, and "disposition" predominates. This observation is not merely a matter of tallying words, but part of a broader conceptual analysis aimed at avoiding the conflation of distinct yet related terms in an effort to trace the development of Ramsey's thought. Since I argue that "habit" and "disposition" are not reducible to one another within his framework, examining how he uses the two terms across his writings is a necessary part of this philosophical analysis. It helps clarify not just what Ramsey thought, but also how he thought it—how his conceptual vocabulary shaped, and was shaped by, his evolving views on belief.

Notably, Ramsey does not explicitly define habits as acquired dispositions. Instead, he incorporates habits into his theory of belief, which is part of his broader theory of probability as a branch of the logic of partial belief and inconclusive arguments.[9] Ramsey requires a theory of belief in order to adopt a quantitative approach to measuring degrees of belief through a "purely psychological method" (Ramsey 1994, 62). In other words, he is concerned with the distinction between stronger and weaker degrees of belief. This is very important because Ramsey's theoretical points depend on this pragmatic need to approach beliefs from a quantitative standpoint.

Rather than viewing degrees of belief as introspectable feelings, Ramsey treats them as causal properties of the belief that determine "the extent to which we are prepared to act on it" (ibid., 65). This is essentially the idea, already developed by Russell (1921/2008), that the differentia of belief lies in its causal efficacy: "how far we should act on these beliefs" (Ramsey 1994, 66). In this framework, beliefs serve as the foundation for potential actions, guiding behavior in hypothetical situations. The strength of a belief is assessed by proposing a bet and determining the lowest odds one

[9] As a reviewer has rightly noted, it might be objected that Ramsey was critical of traditional definitional approaches in philosophy, particularly those that seek necessary and sufficient conditions for the application of concepts. In his essay *Philosophy* (Ramsey 1994, 1-7), Ramsey argues for a more pragmatic and elucidatory role for philosophical analysis. I fully agree with this interpretation. My claim that Ramsey "does not explicitly define" habit as an acquired disposition should not be taken to imply that he fails to provide such a definition in the classical sense. Rather, I mean that he does not offer a formal or reductive account of habit in terms of disposition. His remarks on habit are better seen as part of a broader, non-reductive conceptual elucidation, consistent with his anti-definitional stance. I aim to show that while Ramsey treats belief in dispositional terms and employs the notion of habit, he does not treat habit and disposition as interchangeable or reducible to one another. In this respect, my approach mirrors Ramsey's own methodological commitment to clarifying our conceptual practices rather than imposing rigid theoretical definitions on them.

would accept. Thus, beliefs are that basis for possible actions; they are ideas that can lead to certain actions under certain circumstances.[10] Given this, the notion of habit is a fundamental component of a psychological approach that, while only an approximation of truth, allows for a theory of quantities that is both general and exact. This theory rests on two key ideas: first, that people act in the way they believe most likely to realize the objects of their desires—so that action is determined by desires, interests, and opinions; and second, that the human mind operates "essentially according to general rules or habits" (ibid., 90).

The notion of disposition appears only once in the aforementioned texts, in the phrase "dispositional belief" (ibid., 68), which is contrasted with actualized belief. A dispositional belief is a belief that guides action in cases where it is relevant, though it is rarely considered, such as the belief that the Earth is round. A belief is actualized when a person thinks of it at a particular moment. Beliefs that are considered the basis for action are dispositional insofar as they guide action when relevant without being explicitly considered by the agent. In this context, "dispositional" means "not actualized"; the contrast is between something potential acting in the background and something occurrent explicitly operating at a conscious level. Not surprisingly, this distinction aligns with the distinction between dispositional states and definite acts of thought established in **OT**, where the notion of disposition plays a central role.

On Truth is a posthumous work based on manuscripts that Ramsey wrote between 1927 and 1929. He intended to develop a comprehensive study of truth and probability, synthesizing his earlier work into a unified whole. On a general level, Ramsey's "truth project" can be outlined in four steps. First, he frames the problem of truth as a question about the meaning of "true" when applied to mental states—not what is true, but what truth is. Second, he defends a redundancy theory of truth based on propositional reference against coherentist and pragmatist views. According to this theory, a belief is true if it is a belief *that p*, and *p*. A belief is false if it is a belief *that p*, and *non-p*.[11] However, Ramsey argues that correspondence alone is insufficient to properly account for what we mean by truth and propositional reference because the correspondence relation is not unique

[10] Like his pragmatist predecessors, Ramsey linked belief to action. In particular, he adopted Alexander Bain's view that a belief involves "acting, or being prepared to act, when the occasion arises" (Bain 1872, 372) and Peirce's view that belief involves "the establishment in our nature of a rule of action, or, say for short, a habit" (Peirce 1930-51, 255).

[11] Here lies a possible misunderstanding of the label "pragmatist". Ramsey actually criticizes a particular pragmatist theory of truth—the idea that a belief is true if it is useful (Ramsey 1991, 17-18). However, his theory of truth can still be called pragmatic. Indeed, according to Tuzet (2020), Ramsey finds it useful to link truth and utility insofar as it captures an important aspect of propositional reference: the belief that A is B is useful if A is B, and it is not useful if A is not B.

in form, that is, it is not always and necessarily a direct correspondence between a mental state and a fact. Given this, Ramsey must clarify what propositional reference means in order to discuss the kind of correspondence involved. Finally, he clarifies the nature of judgment to defend the idea that correspondence should not be viewed as a relation between judgments and facts, since the objects of beliefs, opinions, and conjectures are not facts but rather propositions or attitudes associated with actions and their utility (Gruber 2022). In other words, it is the propositional content that guides behavior.

Within this project, Ramsey employs the concept of disposition to address the problem of propositional reference and the definition of judgment. He defines judgment as an *act* of thought with propositional reference and an affirmative character. To develop this idea, however, he first distinguishes between two kinds of states of mind: dispositions and acts of thought. Mental states such as knowledge, belief, and opinion are *dispositional* in that they have dispositional characteristics, that is, they exhibit qualities of disposition or character in the ordinary sense, applying to both mental states and material substances. For instance, if I say that Paul knows the date of the Norman Conquest, I do not mean that he is thinking "1066" at the moment of speaking, but rather that he would be able to provide the answer when asked. Similarly, a man can be called brave or irascible without implying that he is currently displaying those qualities. Likewise, when we say that a poker is strong, we mean that it can withstand a considerable strain without breaking, but the poker in question may never be subjected to such a strain. In this regard, Ramsey states that

> A belief is a disposition not only to make a certain kind of judgment on suitable occasions but also to behave in certain ways in pursuing the pursuit of certain ends. (Ramsey 1991, 99)

Dispositional beliefs generate judgments and guide actions based on the principle that individuals act in ways that would yield the most satisfactory consequences if their beliefs were true. In contrast, "'thinking', as in 'I was just thinking it would snow tomorrow' (…) 'judging', 'inferring', 'asserting', 'perceiving', 'discovering', and 'learning' all refer to definite acts [of thought]" (ibid., 45)—real events and dateable acts of mind that can be expressed by dispositions because they manifest underlying dispositions.

3. Against conceptual reduction: Definition*s*

So what are habits and dispositions? The second key difference concerns their definitions. I argue that, broadly speaking, Ramsey conceptualizes mental habits as procedures—methods of thought, laws, or rules—whereas he conceptualizes dispositions as tendencies and—more generally—as a logical category to which the concept of habit belongs. However, a precise characterization of the nature of dispositions remains an open question. Rather than being a rule or law, Ramsey seems to suggest that the term "disposition" refers to an entity, yet he does not clarify whether it is a mere logical construct or the referent of a hypothetical entity. I will first discuss the conceptualization of habit before turning to the notion of disposition. Finally, I will distinguish between two kinds of dispositionalism to support the idea that habits are dispositional without necessarily being dispositions in the metaphysical sense.

3.1 Habits as methods of thought

In **TP**, Ramsey defines habit as "simply (a) rule or law of behaviour, including instinct" (1994, 90). A habit, understood as a rule, can be either acquired or innate. According to Ramsey, the key feature of habit is its role in regulating processes of thought, regardless of its origin.[12] Consequently, he sees no fundamental distinction between acquired habits (or rules) and innate rules (or instincts), as both function as principles of action. This is not an obvious way to conceptualize habit. Indeed, one might emphasize acquisition instead, distinguishing between learned habits and innate instincts.[13] However, Ramsey does not distinguish between habit and instinct for a pragmatic reason: he is concerned with a particular kind of habit—namely, habits of thought, such as habits of inference, observation, memory, induction, and doubt. As we have seen, Ramsey aims to defend a

[12] One might object that attributing the central role of regulating thought to habit stands in tension with Ramsey's assertion that habit "telescopes" thought away. However, this apparent tension can be resolved by distinguishing between the nature of habit and its functioning. On Ramsey's account, the regulation of thought is indeed a central feature of habit's nature, insofar as acquisition is not considered an essential conceptual component. That is, for Ramsey, a habit need not be acquired in order to function as a rule or principle governing action and thought. The notion of "telescoping" thought away should not be conflated with the regulation of thought *per se*; rather, it refers to the process of habit formation—specifically, the way habitual behavior bypasses the intermediate stage of explicit judgment and proceeds in an immediate and unreflective manner. I am grateful to one of the anonymous reviewers for drawing attention to this point.

[13] For instance, habits and instincts are rigidly separated in Baldwin's (1901) *Dictionary of Philosophy and Psychology*: a habit is defined as an individually acquired function, while a custom is defined as a widespread, habitual manner of acting in society that is not physically inherited—when it is, then it is defined as an instinct. Similarly, Dewey defines habit as a "kind of human activity which is influenced by prior activity" (2023, 20). Although Dewey does not draw a rigid distinction between habit and instinct in conduct, he maintains that instinct is prior to habit in individuals' lives, assuming that habits are learned and instincts are not.

causal theory of belief and holds that a psychological theory must be assumed to measure a rational agent's degrees of belief. This psychological framework posits that the human mind operates according to general rules. Thus, Ramsey's focus is on laws, methods, and procedures of thought—such as the habit of proceeding from the opinion that a toadstool is yellow to the opinion that it is unwholesome. It is for this reason that he focuses on habits as principles of action, irrespective of their origin. As he states, "whenever I make an inference, I do so according to some rule or habit" (ibid., 91), for a process of thought that does not proceed according to some rule is merely a random sequence of ideas.

Given this, mental habits exhibit at least two key features: (1) they entertain or produce varying degrees of beliefs and opinions, and (2) they may be useful or useless depending on how closely the degree of belief they produce aligns with the actual proportion in which they lead to truth (ibid., 92). This means that, according to Ramsey, mental habits can either lead to truth or diverge from it. For this reason, they can be evaluated and judged in a pragmatic way, that is, by whether they work or not—whether the resulting opinions are for the most part true, or more often true than those resulting from alternative habits. Consequently, we can only praise or blame opinions indirectly, insofar as we praise or blame the habits that produce them. All beliefs involve habits, that is, we deduce from them and act accordingly. Indeed, for Ramsey, the question of the ideal is nothing more than the question of what habits would best serve the human mind, given that habits lead to opinions and are more or less conducive to the truth. The view that mental habits are procedures and methods of thought evaluated pragmatically is further reinforced by Ramsey's treatment of induction, variable hypotheticals, and knowledge.

First, consider the case of induction. According to Ramsey, induction is not just a mental habit, but a *good* mental habit, because it generally leads to true opinions and is regarded as a reliable process. Indeed, we all agree that a man who did not make inductions would be unreasonable. In this context, however, "reasonable" and "unreasonable" do not mean, respectively, in accordance with and against formal logic, but rather possessing a good and useful habit or thought procedure—one that increases the likelihood of forming true beliefs (ibid., 93). It is important to note that Ramsey's perspective is general: he is not claiming that every induction leads to truth—counterexamples abound—but rather that induction is "for the most part" a procedure that leads to truth. In other words, induction is generally a truth-conducive method.

Second, consider the case of variable hypotheticals, such as "Arsenic is poisonous" and "All men are mortal". These are general propositions in which the variable remains unrestricted, making them open generalizations. In **GPC**, Ramsey analyzes variable hypotheticals in terms of mental habits, countering the view that they are conjunctions (ibid., 148-149).[14] Their defining characteristic is that they express an inference that we are prepared to make at any time, rather than a belief of the primary sort; in other words, they express a dispositional belief. For example, to believe that all men are mortal is partly to say so and partly to believe, in regard to any *x* that turns up, that if *x* is a man, then he is mortal.

Ramsey defines variable hypotheticals as encapsulations of rules of judgment that form the system with which we meet the future. Yet, these rules are mental habits and, as we have seen, all beliefs involve habits. In this context, habits are rules that make up a system with which we meet the future—rules that enable to state, "If I meet *q*, then I have to treat it as *p*"—and a variable hypothetical is a trustworthy rule, a useful and working mental habit. In other words, habits are rules and variable hypotheticals are good and trustworthy rules—that is, good mental habits—because they generally lead to true beliefs.[15] Indeed, in **K** (ibid., 110), Ramsey uses the concept of a variable hypothetical to clarify what "reliable way" means in the definition of knowledge as a belief which is true, certain, and formed in a reliable way. Here, certainty is psychological rather than epistemic. However, upon reflection, we realize that we can only have certainty if we regard our way as reliable. This, in turn, involves formulating as a variable hypothetical the habit of following the way because it is considered a good or useful habit. In all these cases, a habit is a "method of thought" (ibid., 94) which is responsible for certain beliefs, opinions, and courses of action. Furthermore, mental habits are not particular procedures fixed through rough repetition; rather, they are *regular*, *structured*, and *shared* thought processes that constitute the implicit mental background of human thought.

Against this conceptualization of habit, one might argue that habits cannot be methods (of thought) because they differ in at least two key respects. First, the concept of method is broader than that of habit. For example, it would be odd to claim that if I have the habit of thinking that a certain wine

[14] This is the view that the universal quantifier, "$\forall x\ \phi(x)$", is a conjunction "$\phi(a) \wedge \phi(b) \wedge \phi(c) \wedge\ldots$" and the existential quantifier, "$\exists x\ \phi(x)$", is a disjunction "$\phi(a) \vee \phi(b) \vee \phi(c) \vee\ldots$". Wittgenstein held this idea in his *Tractatus logico-philosophicus* (1961). For a detailed discussion of Ramsey's critique of this theory and its influence on Wittgenstein's later philosophy, see Marion (2012).

[15] This is not to deny the existence of bad habits or untrustworthy variable hypotheticals; rather, the point is that the relevant criterion of demarcation is practical, not formal. Habits may be either good or bad, depending on whether the opinions to which they give rise are mostly true. A variable hypothetical qualifies as a good habit not by virtue of its logical form, but by the role it plays within a given system of thought. The proposition "All men are immortal", for example, is formally a general proposition and thus a variable hypothetical for Ramsey. However, it is not trustworthy, insofar as it is not among the variable hypotheticals that constitute the system "with which we meet the future". I am grateful to one of the anonymous reviewers for highlighting this issue.

is red because it is from Tuscany, I thereby possess a corresponding method of thought. Second, the concept of method is often used normatively, that is: it does not describe how people think and act, but how they *ought to* think and act. In contrast, the concept of habit appears to be purely descriptive, as a habit of thought describes how someone thinks, rather than how she ought to think.[16] *In principle*, however, the concept of habit is not incompatible with generality and normativity. Ramsey's notion of habit refers to a kind of habit that is collective rather than individual, and normative rather than merely descriptive. Of course, this does not mean that all habits are collective and normative, but rather that the variety of habitual behavior is more complex than what is commonly assumed.

As far as generality is concerned, a habit is a collective and general method of thought because it is a stable, entrenched practice that individuals learn due to belonging to a particular form of life. Some collective habits, or customs, form the system into which people are born and raised, making the environment already deeply habitualized in this respect.[17] This is a pragmatist theme, emphasizing that custom has a cumulative dimension that should not be forgotten. As Dewey observes:

> "[T]he activities of the group are already there, and some assimilation of his own acts to their pattern is a prerequisite of (…) having any part in what is going on". (2023, 33)

We do not need to "build private roads to travel upon", but it is convenient and "natural to use the roads that are already there" (ibid.).

Similarly, Young describes custom as "both architect and policeman of society" (1988, 99), because, even though there is always room for variation in principle, custom is the factor involved in the constitution of the regularity of society. An individual may be said to have a habit of induction, but induction itself is not a private road that must be built anew every time; rather, it is one of the paths that constitute the form of life to which the individual belongs. This form of life is shared with other individuals and is transmitted through formal and informal processes of socialization, education, and continuous interaction with the broader environment.

As regards normativity, a mental habit can be considered normative if it functions as a rule or standard for thought and action. In this sense, it not

[16] I thank Giovanni Tuzet for this important point.

[17] The term "custom" is used by Dewey precisely to grasp the collective dimension of habit. Customs are "wide-spread uniformities of habit" (Dewey 2023, 33).

only describes how an individual thinks, but it is also one of the elements in virtue of which an individual acts and thinks in a particular way. More precisely, we might observe that when we say that a habit describes how people think, we are not necessarily divorcing habit from normativity: a habit may describe the methods and principles according to which people think and act.

This is particularly relevant if we think about habits of thought.[18] Indeed, in Ramsey's time, this notion was not uncommon; Ludwig Wittgenstein and Ludwig Boltzmann also employed it (Preston 2022). Boltzmann (1974), for instance, argued that many "illusory" philosophical problems stemmed from deceptive habits of thought—that is, habits useful in specific contexts but misleading when overextended. They overshoot the mark, so to speak. Consider philosophical reductionism: the activity of dismantling "concepts into simpler elements and explain[ing] phenomena by means of laws we know already" is both useful and necessary, but it

> becomes so much a habit as to produce the compelling [but misleading] appearance that the simplest concept themselves must be dismantled into their elements and the elementary laws reduced to even simpler ones. (Ibid., 137)

In a similar vein, Wittgenstein noted that "by the force of habit, we are [so] accustomed to calming our mental anxieties by reducing certain propositions to others that are more fundamental" that we tend to adopt this remedy even when it is practically useless. For example, we do this when our anxiety arises from a lack of clarity regarding the grammatical connections in certain linguistic domains (Baker et al. 2003). We may become so accustomed to particular procedures and methods of thought that Wittgenstein defines even the laws of logic as expressions of "thinking habits" and "the habit of thinking" (Wittgenstein 1978, §131). The latter shows "what human beings call 'thinking'", but the former shows "how human beings think" (ibid.) because such a habit is at the basis of action and thought: "thanks to custom, particular forms become paradigms; they acquire the force of a law" (Wittgenstein 1980, I §343).[19]

[18] I am not defending the idea of a rigid distinction between *purely* mental habits and *purely* bodily one. The very notion of habit allows us to bypass mind-body dualism. Nevertheless, I conceptually distinguish between them, since not all habits are paradigmatically manifested in actions and movements. In other words, not all habits are motor habits: some take the form of implicit, embodied practices and procedures that guide our inquiry. Some of them are also embedded in institutionalized practices, such as the habit of induction.

[19] Dewey (2023, 33-35) too acknowledges the normative import of collective customs that become laws, regulative patterns, and standards for individual conduct.

3.2 Dispositionalism

According to Ramsey, dispositional states govern our actions and give rise to corresponding judgments, which are themselves acts. As a category, they differ from occurrent states of mind in that: (1) they refer to the persistent background of the mind, rather than to discrete spatio-temporal acts of thought; (2) they are potential, that is, they would manifest under the right conditions, but they may also remain latent; (3) they state a purely hypothetical fact, namely what a person would think, say, or do; and (4) they are primarily manifested in action rather than thought (Ramsey 1991, 43-45). In this regard, Ramsey treats dispositionality as a logical-grammatical category to which certain psychological concepts belong.

This grammatical approach to dispositionality reappears in Wittgenstein's *Remarks on the Philosophy of Psychology* (1980, II §§ 43, 45, 178, 243), where he distinguishes between the logical category of dispositions and that of states of consciousness. The distinction serves to clarify the difference between concepts such as understanding, meaning, intending, and believing, on the one hand, and concepts such as feeling pain, perceiving something, and seeing an image, on the other. For Wittgenstein, this distinction is purely methodological; he employs it as a conceptual tool to reject the idea that psychological concepts refer to the inner, private states, processes, or experiences [*Erlebnisse*] of subjects.[20]

However, unlike Wittgenstein, Ramsey also maintains that (6) dispositional states depend on non-dispositional states (1991, 44), and that (7) they explain a kind of immediate (conditioned-reflex) action or response to stimuli (ibid.,50). These two claims imply a shift toward using the concept of disposition to refer to a particular entity that is hypostatized, ascribed to the agent, and invoked to causally explain other mental states and processes. In doing so, Ramsey partially moves toward a kind of metaphysical dispositionalism that postulates the existence of dispositional entities rather than using dispositionality purely as a logical-grammatical category. In contemporary debates on dispositions, this metaphysical shift is assumed by both dispositional realism and categoricalism, though they differ in how they characterize the nature of dispositional entities. Dispositional realists (e.g. Mellor 1974; Mumford 2003) argue that dispositions possess ontological autonomy, whereas categoricalists claim that all dispositional properties ultimately depend on some underlying non-dispositional properties of their bearers, such as their molecular structure or biological system (Armstrong 1997).

[20] For a detailed analysis of Wittgenstein's use of the concept of disposition, see Morelli (2024).

Ramsey *partly* leans toward metaphysical dispositionalism because he appears to endorse classical categoricalism in (6). However, he endorses a specific version of categoricalism. Traditionally, categoricalism has been understood as both an ontological thesis about the nature of dispositional entities and a semantic thesis about dispositional terms, where non-dispositional properties are the referents of the dispositional expressions.

However, Ramsey does not formulate categoricalism from a linguistic perspective. Instead, he appears to support the idea that dispositional characteristics depend on some positive characteristics, because he endorses the analogy between mental dispositions and the physical dispositions of inert matter. For example, in the case of a poker, we suppose that its strength—a dispositional characteristic—depends on the non-dispositional properties of its constituent particles. Similarly, in the case of knowledge, which is a dispositional state, we suppose that it depends on some "arrangement, trace or record" (Ramsey 1991, 44) formed in the mind or brain through learning and retained until forgotten.

Yet, according to Ramsey, these positive characteristics are not the referents of dispositional terms. We can discuss dispositional characteristics meaningfully without identifying their categorical basis because we only need to know the kinds of actions, thoughts, and reactions they are expected to produce. In this sense, from an ontological perspective, "the problem of their status is very analogous to that of the unobservable entities in physics" (ibid., 101). A disposition is thus conceptualized as an unobservable property, entity, or character inferred from external behavior and used to explain specific actions and thoughts, even when knowledge of these entities is lacking—whether they be material particles, mental traces, or brain processes. Nevertheless, we do explain behavior in dispositional terms, regardless of our knowledge of its supposed categorical basis. Therefore, positive characteristics are posited to account for the latent nature of dispositions—the fact that they persist even when not concretely manifested. However, we could also treat dispositions as mere "logical constructions" without undermining the validity of dispositional discourse and explanation.[21]

[21] This perspective is drawn from Broad (1925), a work explicitly cited by Ramsey in **OT** (1991, 44, footnote 1). Broad uses the notion of disposition alongside that of mental trace to reconstruct the debate on mnemic persistents—mental entities produced by experience that persist and give rise to new states of mind or modify existing ones when triggered by stimuli. These entities were postulated to explain forms of behavior learned from past behavior, despite the large time gap between the two. From a metaphysical perspective, Broad carefully weighs the advantages and disadvantages of two theories concerning the nature of these mental persistents: trace theory and mnemic causation. However, he argues that there is no need to choose between them, since we use these notions to predict, control, and explain mental events but know nothing about these entities in detail. We know what they are only through their effects, and we know it independently of the particular theory on their nature we choose to adopt.

On this basis, I conclude that Ramsey does not view dispositions as rules or laws. They are not procedures of thought, like mental habits. Rather, dispositionality is, first and foremost, a logical category to which mental states such as believing, knowing, and intending belong. Second, dispositions are conceptualized as tendencies or propensities toward particular thoughts and actions engendered by certain mental habits. This suggests that the notion of habit is itself dispositional from a grammatical standpoint, but not in a metaphysical sense, as habits are not metaphysically reduced to dispositional entities.

4. Belief, judgment, and knowledge

In this section, I will elaborate on the idea that thought habits are dispositional insofar as they produce beliefs and opinions that, in turn, dispose an agent to think and act in a particular manner. However, they should not be conflated with dispositions themselves, because Ramsey does not reduce them to dispositional entities. Additionally, I will examine the aforementioned point (7), which concerns the use of dispositional language to explain immediate and conditioned-reflex action in response to stimuli.

As we have seen, Ramsey distinguishes between two logical categories: dispositions and definite acts of thought. While belief falls under the dispositional category, judgment belongs to the occurrent category. A judgment is (a) an act of thought that (b) has propositional reference *and* (c) an affirmative character. Propositional reference is the *aboutness* of a mental state—its being "that something is so and so" (Ramsey 1991, 7). Consequently, a judgment (a) is not a dispositional belief, (b) is an occurrence of thinking that something has a particular property, and (c) is not a state of doubt or wondering. In this sense, we could say that "judgment" is a more precise word for actual belief, as opposed to dispositional belief.

From this perspective, Ramsey's inquiry into the concept of truth offers insight into the relationship between beliefs, habits, dispositions, judgments, and knowledge. The idea is that, on the one hand, a belief—dispositional in nature—disposes the agent to make particular judgments, which can, in turn, be seen as manifestations of that belief. On the other hand, judgments, as acts, can give raise to particular dispositions. In general, judgments are particular acts of thought with propositional reference and an affirmative character. However, they vary depending on the kind of belief they manifest or the type of disposition they produce: there are judgments "which are cases of knowing" and judgments "which are not knowledge". The latter are then called "opinions" (ibid., 55). In this context, Ramsey seems to imply that knowledge is a specific kind of judgment—i.e., that which is nearly always true, certain, and justified—yet this clashes with the claim that the concept of knowledge belongs to the dispositional category, since judgment is classified as an act of thought. Ramsey is perfectly aware of this, and he addresses the issue right at the beginning of the chapter on knowledge and opinion in **OT**. He starts by stating that "judgment in the wide sense in which we use the term was held (…) to comprehend two essentially different processes, knowledge and opinion" (ibid.). His next step in this chapter is to examine this distinction. At the same time, though, Ramsey admits that "the words knowledge and opinion (…) are most commonly used not of judgements but of dispositions". However,

> since the distinction which we are investigating is primarily one between different kinds of judgements, (which can be extended to the dispositions arising from or manifested by these two kinds of judgements) we shall use the words knowledge and opinion in the present chapter to mean judgments and not dispositions. (Ibid.)

Conceptually, knowledge belongs to the category of the dispositional. Yet, in this *specific* context, Ramsey treats knowledge as a judgment because he wants to "examine the meaning and validity" of the ordinary distinction between knowledge and opinion—that is, the view that knowledge involves certainty, whereas opinion carries some degree of uncertainty. In particular, treating knowledge as a specific kind of judgment enables Ramsey to claim that knowledge and opinion are not two different classes with different propositional references, but rather different species of the same genus—a false judgment is not considered knowledge, yet it remains a judgment nonetheless. At the same time, we have seen that all beliefs involve habits, that is, we deduce from them and we act accordingly in a certain way, meaning that the dispositions manifested in particular acts of thought are engendered by certain habits of mind. To develop this point further, we must examine the notion of judgment more closely.

Ramsey's concept of judgment presupposes no process of reflection or weighing of evidence: a judgment can be a reasoned conclusion, a guess, a prejudice, a memory, or a presentiment, provided it has propositional reference and an affirmative character. However, judgment appears to be essentially *mediated* and to require a *thought* as a response. Indeed, Ramsey does not classify immediate action, where no mental intermediary is involved, as a case of judgment. For example, uttering "It's a fly" is a judgment, whereas swatting the fly without saying anything is not a judgment, even indirectly. Similarly, consider the case of seeing an apple. I may conjure up images of the taste of apples or the word "apple", or I may *directly* experience significant bodily changes or actions without articulating any thoughts. The former case is a judgment, the latter is not. The latter can be understood as an action done out of habit, that is, triggered directly by a certain stimulus. Indeed, Ramsey defines a habitual action as one that is *directly* triggered by an external stimulus, because "habit or instinct has made the intermediate stage of judgement disappear" (ibid., 51). Accordingly, he refers to acting out of an old habit as acting without explicitly thinking about it, in the sense of engaging in a definite act of thought or forming a mental image *beforehand*. For this reason, habitual action is a manifestation of a dispositional belief function *without* judgment.

In this respect, Ramsey appears to endorse what is now called "the received view" on automaticity (Douskos 2013), which is the idea that automaticity is an essential feature of habitual acts involving the absence of deliberation and intention. Habitual acts are *directly* triggered by a stimulus and are direct responses to circumstances. Now, one way to explain the motivational force of habit—the idea that something is done out of habit—is to say that habit disposes the agent in a certain way. Ramsey's use of the notion of disposition goes in this direction, because reference to dispositions is a way to characterize the tendency to act, or habit's motivational pull. For this reason, habit can be said to be dispositional in nature from a logical point of view. Habit has a dispositional profile, so to speak.

Given this, is habit dispositional in a metaphysical sense too? In other words, is habit itself a disposition? As we have seen, Ramsey leaves open the question of whether the term "disposition" is purely a logical construct or denotes an actual dispositional entity, such as a mental trace or record. Regardless of this metaphysical issue, he does not reduce habit to disposition. He does not conceptualize habit as an internal matrix or source of actions and thoughts. Instead, he treats habit as having a dispositional character from a logical standpoint, framing it as a tendency toward certain actions—an inherent motivational pull. Thus, while habit has a dispositional profile, its metaphysical nature remains an open question within Ramsey's framework.

Before turning to the conclusion, I will briefly address a contested issue in the literature on Ramsey's philosophy concerning the concept of judgment—an issue which, I argue, can be clarified by the conceptual analysis of habit and disposition developed so far.

4.1 A controversy in the literature

As we have seen, in **OT**, Ramsey asserts that the term "judgment" refers to specific acts of thought, as opposed to dispositions. He distinguishes between judgments that constitute knowledge and judgments that constitute opinions. In **GPC**, he introduces a further distinction within his account of general propositions, distinguishing between judgments and rules for judging. Ramsey contends that "variable hypotheticals are not judgements but rules for judging 'If I meet a ϕ, I shall regard it as a ψ'" (Ramsey 1994, 149). However, scholarly interpretations of this distinction remain divided.

Richard Holton and Huw Price (2003) construe the distinction as one between beliefs of the primary sort—genuine judgments—and universal beliefs, or variable hypotheticals—rules for judging. Yet, they ultimately reject the distinction's significance, arguing that under Ramsey's broader thesis that all beliefs are dispositions, there is no functional or hierarchical difference between the two. While Ramsey characterizes primary beliefs as "a map of neighbouring space by which we steer" (1994, 146), Holton and Price counter that "surely the use of maps is itself dispositional" (2003, 326).

In contrast, Marouzi (2024, 19) contends that this distinction represents a difference in kind. On this reading, Ramsey's central claim is that rules for judging are not propositions, but rather cognitive attitudes that are irreducible to judgments. Drawing on Hugh Mellor, Misak similarly maintains that "singular beliefs, general beliefs, and conditional beliefs are all dispositions to behave, but (…) they correspond to different kinds of dispositions" (2016, 196): beliefs of the primary sort are dispositions to direct action, while open generalizations involve dispositions to acquire other beliefs. My own position aligns with those proposed by both Misak and Marouzi.

First, it is essential to contextualize Ramsey's terminological and metaphorical choices. As we have seen with the term "knowledge", Ramsey often adapts terminology, employing it in broader or narrower senses depending on the problem at hand. Holton and Price (2003) interpret the distinction between judgments and rules for judging primarily within the framework of Ramsey's remarks on infinity. Notably, Ramsey rejects the view that variable hypotheticals are conjunctions. His arguments here rely on the limitations imposed by infinity: a variable hypothetical "cannot be written out" as a conjunction; its application in class-thinking is only valid for finite classes; "it always goes beyond what we know or want"; and its certainty pertains to a particular instance, not to "an infinite number which we never use, and of which we couldn't be certain at all" (Ramsey 1994, 145-146). However, in assessing the significance of the distinction in **GPC**, one must consider the broader context of Ramsey's "human logic" and his commitment to a "realistic spirit". In this respect, I concur with Mathieu Marion (2012, 17), who argues that the introduction of variable hypotheticals is a pragmatic argument not limited to the infinite case at all. Ramsey rejects the treatment of variable hypotheticals as propositions conceived in the Tractarian sense, namely as pictures of facts:

> Variable hypotheticals have formal analogies to other propositions which make us take them sometimes as facts about universals, sometimes as infinite conjunctions. The analogies are misleading. (Ramsey 1994, 160)

Second, Ramsey's characterization of a belief of the primary sort as "a map of neighbouring space by which we steer" is consistent with his treatment of judgment in **OT**, provided we distinguish between habits as methods or procedures and dispositions as tendencies. As we have seen, judgments are definite acts of thought for Ramsey—occurrent mental states that express belief about a specific situation or fact. By contrast, rules for judging are dispositional or procedural: they guide the formation of future judgments. They are not judgments themselves, but rather frameworks that shape subsequent cognitive responses. Judgments, thus conceived, are manifestations of beliefs (understood dispositionally) that can, in turn, generate further dispositions. Therefore, the distinction between definite acts of thought and dispositions is logical, but not formal—in essence, it is functional. In this regard, it is closely akin to Wittgenstein's distinction between grammatical and factual propositions.

Ramsey's characterization of variable hypotheticals as rules emerges as a more realistic alternative to their analysis as conjunctions. This shift is not merely semantic but epistemological: it marks a move away from an analysis based solely on syntactic form and toward one that considers the various cognitive attitudes we might adopt toward a proposition. On this account, a belief of the primary sort—a genuine judgment—is also dispositional. While Holton and Price (2003) rightly note that using a map is itself dispositional, the map's capacity to dispose us toward certain actions partly stems from its representational content (Marouzi 2024, 18). By contrast, a rule does not function by representing how the world is; rather, it expresses a habit of thought. Agents may be disposed in various ways, including through the acquisition of such cognitive habits.

5. Concluding remarks

My argument provides a contribution on two levels: the interpretation of Ramsey's philosophy and the conceptualization of habit and disposition.

On the first level, I have argued that habit and disposition are distinct concepts in Ramsey's philosophy and are not *conceptually* reducible to one another. They share three key features: 1. they are principles of action; 2. they are manifested in particular acts; and 3. they explain immediate action. However, they are used in different philosophical contexts for distinct theoretical purposes and are conceptualized differently. I have differentiated between logical-grammatical and metaphysical dispositionalism, contending that habit is dispositional in nature and should be understood as part of the conceptual framework through which we describe tendencies to act, rather than as an occurrent mental state. Yet, habit itself is not a disposition. Ramsey conceptualizes mental habits as methods of thought—deeply internalized procedures underlying certain actions and thoughts—whereas dispositions emerge from habits as tendencies explaining their motivational force, even though their ontological status remains an open question. This distinction suggests that habits and dispositions are not only conceptually but also metaphysically irreducible to one another.

I have also argued that the distinction between habits and dispositions also elucidates Ramsey's distinction between judgments and rules for judging, especially when viewed through the lens of his broader pragmatic orientation. Judgments, as definite acts of thought, and rules for judging, as procedural habits, play distinct roles in our cognitive life. They differ not merely in grammatical form, but in their functional orientation toward future reasoning and action. This interpretation establishes Ramsey's philosophy as one that resists rigid categorization, highlighting instead the dynamic interplay between acts, attitudes, and their practical contexts. Recognizing this distinction clarifies Ramsey's conceptual framework and enriches our broader understanding of belief, judgment, and the dispositional architecture of thought.

On the second level, I have argued that Ramsey's notion of mental habit is both general and normative. It serves as a shared standard of action and thought, constituting the framework of a particular form of life. Moreover, Ramsey develops the idea that habits function as dispositions in the sense that specific habits of thought lead to beliefs, which in turn predispose an agent to act in a certain way. This provides an account of immediate action that bypasses judgment and intellectual mediation. While this view implies a narrow understanding of automaticity as the absence of deliberation, it also frames the relationship between habit and disposition without necessarily postulating an additional dispositional entity to explain habitual action causally.

Acknowledgments

This paper is the outcome of research conducted within the framework of the PRIN 2022 project "Habits in (Time of) Crisis" (HiToC). Conceptual Tools for Dealing with Disruptive Events", funded by the European Union—NextGenerationEU—under the National Recovery and Resilience Plan (PNRR), Mission 4, Component 2, Investment 1.1 "Fund for the National Research Programme and Projects of Significant National Interest (PRIN)". I am grateful to Professor Giovanni Tuzet for reading earlier drafts of this manuscript and for generously providing valuable suggestions and feedback. I also wish to thank the two anonymous reviewers of this journal for their critical and constructive comments on earlier versions of the paper.

REFERENCES

Armstrong, David M. 1997. *A World of States of Affairs.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511583308

Baker, Gordon P. ed. 2003. *The Voices of Wittgenstein—The Vienna Circle, Ludwig Wittgenstein and Friedrich Waismann*. London: Routledge. https://doi.org/10.4324/9780203412022

Bain, Alexander. 1872. *Mental and Moral Science*. London: Longman, Green, and Co.

Baldwin, James M. ed. 1901. *Dictionary of Philosophy and Psychology*. New York: The Macmillan Company.

Boltzmann, Ludwig. 1974. *Theoretical Physics and Philosophical Problems: Selected Writings*. Dordrecht: D. Reidel.

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.

Broad, Charlie D. 1925. *The Mind and its Place in Nature*. New York: Harcourt, Brace & Company, INC.

Chiaradonna, Riccardo, and Flavia Farina. 2022. "Aristotle on (second) Nature, Habit and Character." In *The Routledge Handbook of Liberal Naturalism*, edited by Mario De Caro and David Macarthur, 7-17. London: Routledge.

Dewey, John. 2023. *Human Nature and Conduct*. Zinc Read.

Douskos, Christos. 2013. “Deliberation and Automaticity in Habitual Acts.” *Ethics in Progress* 9 (1): 25-43. https://doi.org/10.14746/eip.2018.1.2

Engel, Pascal. 2005. “Belief as a Disposition to Act: Variations on a Pragmatist Theme.” *Cognitio* 6 (2): 167-185.

Gruber, Monika. 2022. “Ramsey’s Theory of Belief.” *European Journal of Pragmatism and American Philosophy* 14 (2): 1-19. https://doi.org/10.4000/ejpap.3018

Holton, Richard, and Huw Price. 2003. “Ramsey on Saying and Whistling: A Discordant Note.” *Noûs* 37 (2): 325–341. https://doi.org/10.1111/1468-0068.00441

Hutto, David, and Ian Robertson. 2021. “Clarifying the Character of Habits.” In *Habits. Pragmatist Approaches from Cognitive Science, Neuroscience, and Social Theory*, edited by Fausto Caruana and Italo Testa, 204-222. Cambridge: Cambridge University Press. http://doi.org/10.1017/9781108682312.010

James, William. 1914. *Habit*. New York: Henry Holt and Company.

Kraemer, Eric R. 1985. “Beliefs, dispositions and demonstratives.” *Australian Journal of Philosophy* 63 (2): 167-176. https://doi.org/10.1080/00048408512341781

Lockwood, Thornton. 2013. “Habituation, Habit and, Character in Aristotle’s Ethics.” In *A History of Habit: From Aristotle to Bourdieu*, edited by Tom Sparrow and Adam Hutchinson, 19-36. Lanham, MD: Lexington Books.

Marion, Mathieu. 2012. “Wittgenstein, Ramsey, and British Pragmatism.” *European Journal for Pragmatism and American Philosophy* 4 (2). https://doi.org/10.4000/ejpap.720

Marouzi, Soroush. 2024. “Frank Ramsey’s Anti-Intellectualism.” *Journal for the History of Analytical Philosophy* 12 (2): 1-31. https://doi.org/10.15173/jhap.v12i2.5469

McKitrick, Jennifer. 2009. “Dispositions, Causes, and Reduction.” In *Dispositions and Causes*, edited by Toby Handfield, 31-64. Oxford: Clarendon Press; https://doi.org/10.1093/oso/9780199558933.003.0002

Mellor, Hugh. 1974. “In Defense of Dispositions.” *The Philosophical Review* 83 (2):157-181. https://doi.org/10.2307/2184136

Merleau-Ponty, Maurice. 2012. *Phenomenology of Perception*. New York: Routledge.

Mills, John A. 1998. *Control. A History of Behavioural Psychology.* New York and London: New York University Press.

Miner, Robert C. 2013. “Aquinas on *Habitus*.” In *A History of Habit*, edited by Tom Sparrow and Adam Hutchinson, 67-87. Lanham: Lexington Books.

Misak, Cheryl. 2016. *Cambridge Pragmatism. From Peirce and James to Ramsey and Wittgenstein*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198712077.001.0001

Misak, Cheryl. 2022. "The Pragmatism of Ramsey and Ryle." In *A Tribute to Ronald de Sousa*, edited by Julien Deonna, Christine Tappolet, and Fabrice Teroni. https://www.unige.ch/cisa/related-sites/ronald-de-sousa/

Miyahara, Katsunori, and Ian Robertson. 2021. "The Pragmatic Intelligence of Habits." *TOPOI* 40: 597-608. https://doi.org/10.1007/s11245-020-09735-w

Morelli, Alice. 2024. *A Wittgensteinian Perspective on Dispositions*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-60506-2

Mumford, Stephen. 2003. *Dispositions*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199259823.002.0001

Peirce, Charles S. 1930-51. *Collected papers of Charles Sanders Peirce*, edited by Burks and Weiss. Cambridge, MA: Harvard University Press.

Preston, John. 2022. "The Idea of a Pseudo-Problem in Mach, Hertz, and Boltzmann." *Journal for General Philosophy of Science* 54: 55–77. https://doi.org/10.1007/s10838-021-09569-z

Ramsey, Frank P. 1991. *On Truth*, edited by N. Rescher and U. Mayer. Dordrecht: Kluwer Academic Publishers.

Ramsey, Frank P. 1994. *Philosophical Papers*, edited by D. H. Mellor. Cambridge: Cambridge University Press.

Ravaisson, Félix. 2008. *Of Habit*. London: Continuum.

Russell, Bertrand. 1921/2008. *The Analysis of Mind*. Sioux Falls: NuVision Publications, LLC.

Tuzet, Giovanni. 2020. "Introduzione. Perchè è meglio credere il vero." In *Sulla verità e scritti* pragmatisti, edited by Giovanni Tuzet, 9-32. Macerata: Quodlibet.

Wittgenstein, Ludwig. 1961. *Tractatus Logico-Philosophicus*, translated by D. F. Pears and B. F. McGuinness. London: Routledge and Kegan Paul.

Wittgenstein, Ludwig. 1978. *Remarks on the Foundations of Mathematics*, translated by G. E. M. Anscombe. Oxford: Basil Blackwell.

Wittgenstein, Ludwig. 1980. *Remarks on the Philosophy of Psychology*. Volumes I and II, translated by G. E. M. Anscombe. Oxford: Basil Blackwell.

Young, Michael. 1988. *The Metronomic Society: Natural Rhythms and Human Timetables*. London: Thames and Hudson.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 2 | 229-251
https://doi.org/10.31820/ejap.21.2.7

WHAT IS A RESPONSE TO WRONGDOING?

Nicolas Nayfeld[1]

[1] Jean Moulin Lyon 3 University, France

Original scientific paper – Received: 10/03/2025 Accepted: 11/08/2025

ABSTRACT

This article starts from the assumption that to properly assess the merits of "response retributivism", we must first clarify the nature of a response to wrongdoing and how it differs from a mere reaction. I propose a communicative account of responses, arguing that a response to wrongdoing is a distinctive form of action motivated by the wrong, whose special feature is that it is addressed to the wrongdoer, has a confrontational-communicative dimension, and needs to be identified by the wrongdoer as a response to his wrong. I argue that this definition allows us to rethink the concept of responsibility, to make progress on the debate whether there is a duty to respond or react to wrongdoing, and to refocus the discussion toward what Strawson calls "reactive practices" as opposed to "reactive attitudes".

Keywords: response retributivism; wrongdoing; communicative account; responsibility; reactive practices.

Correspondence: nicolas.nayfeld@univ-lyon3.fr

Introduction

In her 2020 article, Leora Dahan Katz defends a form a retributivism called "response retributivism" and argues that

> [w]hile moral norms generate primary duties to act and refrain from acting in particular ways (primary duties), (…) such norms also generate secondary duties to *react* and *respond* to violations of primary duties in particular ways. (Dahan Katz 2020, 2)

This theory is refreshing because it retains the deontological core of retributivism while expanding the focus beyond punishment, suggesting that punishment is only one possible response to wrongdoing among others.

However, the problem with this theory is that it relies on a fundamental ambiguity as the quote above shows: it remains unclear whether Katz's account implies a duty to *respond* to wrongdoing or a duty to *react* to wrongdoing, or both. This initial problem is compounded by a second: the distinction between responding and reacting, as such, has been neglected by philosophers. They generally do not define the concept of response, taking it for granted and often equating it with Strawsonian "reactive attitudes".[1] Mainstream psychology, for its part, has seized on this distinction, but the way it draws it seems implausible to me: reactions, it is said, are impulsive and unthinking, while responses are thoughtful and deliberate.

This paper has three related aims. The first and main objective is to clarify this distinction and to offer what can be called a communicative account of responses insofar as it shares much with communicative theories of punishment (particularly Duff's) and censure theories (especially Narayan's)—though it extends their insights to responses more broadly, beyond punishment and censure. I should warn from the start that I will not try to determine the necessary and sufficient conditions for a response/reaction to occur. Rather, I will attempt to point out the most salient features.

The second objective is to show that such clarification allows us to make significant progress on the debate whether there is a duty to react or respond to wrongdoing.

[1] For example, although her book is entitled *Forgiveness and Retribution: Responding to Wrongdoing*, Holmgren (2012) does not, to my knowledge, say what it means to respond to wrongdoing.

The third objective is to show that such clarification can help understand what Strawson means by “reactive practices”, a surprisingly understudied notion (see Metz 2008, for an exception), in contrast to “reactive attitudes” on which much ink has already been spilled.

I proceed in six steps. First, I introduce three cases (and variations of each) to make the distinction between not reacting, reacting, and responding intuitive. These cases (“the battered woman”, “sexual harassment on the subway”, “the unruly young child”) serve as a basis for the entire analysis. Second, I identify commonalities between reactions and responses, emphasizing that both involve behavioral action and stand in opposition to inaction, which means that neither letting go nor mere emotions count as reactions or responses within this framework. Third, I explore the differences between reactions and responses, focusing primarily on the specificity of responses, the crucial difference being that responses are *addressed* to the wrongdoer. I explain why this notion of “addressing” is complex and requires careful unpacking. Fourth, I examine the relationship between responses and responsibility on the one hand, and the specificity of penal responses on the other. Fifth, I address a potential objection: that my analysis might overlap with Strawson’s distinction between the objective attitude and the participant attitude or between objective attitudes and reactive attitudes. I argue that this objection is misplaced; rather, a more apt comparison can be drawn between what I call “responses” and what Strawson calls “(reactive) practices” in the opening and closing sections of “Freedom and Resentment”. Finally, I explore the normative implications of this conceptual framework and argue against the idea of a duty to respond to wrongdoing.

1. Three cases

I will start with three cases (and their variations) that vividly illustrate the distinction I want to draw between reacting to a wrong and responding to it. I do not think that the distinction I am going to propose between response and reaction corresponds perfectly to ordinary usage. The distinction between response and reaction in ordinary language is not so clear-cut, and the two terms are sometimes used interchangeably. However, I believe that the distinction I want to draw is *fairly intuitive* in French (my native language) and in English (as native speakers have confirmed to me), that it *corresponds partially to ordinary usage*, and therefore *is not entirely stipulative*. To be clear, in this article, I am not interested in ordinary language as such, but rather in the contrast between different types of situation or action (labelled A series, B series and C series below) that exhibit distinctive features.[2] In fact, it often happens that valuable philosophical distinctions overlap only partially with ordinary language (see for instance Kant's distinction between belief, knowledge and opinion in his *Critique of Pure Reason*).[3]

> Case 1: the battered woman
>
> 1.A. A woman who has faced domestic abuse for years remains silent, refraining from any action out of fear of her husband. She bears her suffering, hoping her ordeal will eventually come to an end.
>
> 1.B. Another woman in a similar situation contacts a domestic violence hotline. She works with volunteers to plan her escape, carefully concealing any evidence of her call to protect herself from her husband's potential retaliation.
>
> 1.C. Another woman, after years of abuse and a final assault, uses a shotgun to kill her husband, then calls the police to confess (a scenario reminiscent of the Jacqueline Sauvage case in 2012).

In 1.A, the abused woman experiences reactive feelings/sentiments/emotions, but in a sense yet to be determined, she neither reacts nor responds to her husband's abuse. In 1.B, the abused woman reacts to her husband's abuse by making a safety plan, though she does not respond directly to it. In 1.C, the abused woman responds violently and irrevocably to the assault she has just suffered, but also to the years of abuse she has endured.

> Case 2: sexual harassment on the subway
>
> 2.A. A bystander witnesses sexual harassment in a subway station but does nothing, held back by fear and discomfort.
>
> 2.B. Another bystander in the same situation discreetly reports the incident to law enforcement, leading to a swift police intervention.
>
> 2.C. A third bystander, upon witnessing the harassment, confronts the perpetrator directly, loudly condemns the behavior, and warns of further consequences if the harassment does not stop.

In 2.A, the bystander fails to react, although the situation does not leave him emotionally indifferent. In 2.B, the bystander reacts by reporting the event through an established channel. In 2.C, the bystander himself provides a response to the harm he witnesses.

[2] For a similar methodological insight, see Hart (2012, v).

[3] I would like to thank reviewer 1 for prompting me to clarify this point.

> Case 3: the unruly young child
>
> 3.A. An 18-month-old child pulls a book from the family bookshelf and starts tearing it. The parents decide to ignore the incident, thinking “it’s no big deal”, “this book is bad anyway”.
>
> 3.B. Different parents in the same situation choose to childproof the bookshelf by moving the books out of reach.
>
> 3.C. Other parents scold the child, explaining that tearing books is unacceptable, attempting to instill a sense of respect for property.

In 3.A, the parents opt for inaction, choosing not to react. In 3.B, the parents react to prevent future incidents but avoid involving the child directly. In 3.C, the parents feel it is appropriate, even necessary, to respond to the incident, and therefore provide (what they believe to be) an educational response to the incident.

Now that we have illustrated the distinction, we can begin to analyze it. I will first explore the commonalities between B series (which I classify as reactions to a wrong) and C series (which I classify I responses to a wrong).

2. Commonalities

Both reactions and responses are motivated and justified by a wrong (an offense, a loss, a harm, a slight, etc.).[4] In B series and C series, the answer to the “why?” question (Anscombe 2000) will mention the wrong. Why did you call the police? *Because he was harassing her.* Why did you start yelling at him? *Because he was harassing her.* A child can push his brother into a puddle because he said something mean to him (response), but also push him “for fun” (not a response) (see Metz 2008, 228).

Both reactions and responses can vary along a spectrum from impulsive and disproportionate to thoughtful and measured. The common assertion, particularly in psychological and self-help literature, that reactions lack any thought while responses are thoughtful is, I argue, mistaken.[5] In anger, a person might respond to an insult with a slap. Conversely, a person might

[4] As my brackets indicate, I use the term “wrong” as a generic term: I am interested in actions that should not have been done and for which someone is accused. Although we can also respond to a benefit, the distinction between responding and reacting does not seem to work as well with benefits.

[5] See https://www.psychologytoday.com/us/blog/the-power-prime/202110/the-difference-between-reacting-and-responding

react with great composure and professionalism to threats from an inpatient in a psychiatric hospital by calmly restraining him.

Both reactions and responses can take place under social pressure. Public figures such as politicians and judges are expected, even required, to react and respond to wrongdoing as quickly as possible. Failure to do so can attract public criticism. This social pressure has arguably intensified with the rise of social media. In fact, Dahan Katz's argument (mentioned in the introduction) in favor of a duty to react and respond illustrates this growing expectation.

Both reactions and responses need not originate with the victims of wrongdoing (if any). They may instead come from third parties, such as bystanders (as in cases 2.B and 2.C) or the state (for instance, when it intervenes to punish criminal wrongdoing). The same is true of linguistic responses (answers): Mary's parents, for example, might respond to Paul's letter by instructing him to cease contact.

Finally, both reactions and responses in B series and C series are behaviors or clusters of behaviors. In case 2.A, although the bystander may experience guilt, he fails to react: as the word suggests, to re-act, you have to act. This is why forgiveness only sometimes qualifies as a response; private forgiveness (e.g., overcoming resentment without any interaction with the offender) is not a response, whereas overt forgiveness (e.g., telling the offender "I forgive you") is. Some may argue that "not resisting" in a situation (such as a bank employee complying calmly with robbers) constitutes a reaction. But this is because "not resisting" in this case involves active behaviors: remaining composed, cooperating, and prioritizing safety, rather than taking no action at all. In fact, both reactions and responses are opposed to inaction in all its forms, whether reflective or unreflective. Deciding to "let go" in the face of wrongdoing, for instance, does not constitute a response; rather, it involves a choice to abstain from responding. Brunning and Milam (2022, 6) give examples of such a choice: a hate-speech victim feels further blame would be pointless and continues on his way; a friend tolerates another's rude behavior to avoid continuous confrontation. Letting go, then, is an option we may choose—or even feel compelled to choose—in the face of wrongdoing, but it does not qualify as a response in the strict sense.

3. Differences

Now let us look at the differences. Responses, as illustrated by the C series above, are always *addressed* to someone, always have an *addressee*—

namely the wrongdoer. If an individual commits mass murder in a school but then commits suicide, there is no way to respond to his crime. However, the authorities can react by imposing stricter gun laws, increasing security in schools, improving emergency evacuation procedures, etc. Reactions need not be addressed to the wrongdoer, though they can address the causes of wrongdoing. Another example: when the identity of the wrongdoer in unknown, it is impossible to respond to his wrongful conduct, for instance by censuring him, "since censure is condemnation addressed to the perpetrator" (Narayan 1993). However, it is perfectly possible to react to his wrongful conduct, for instance by expressing disapproval of it, by denouncing and disavowing it (Narayan 1993).[6]

This means that responses inherently possess a dialogical, face-to-face, "confrontational"[7] (Shoemaker 2015, 112) quality, which is absent in the B series above (reactions). When responding to a wrong, I engage the wrongdoer in a conversation.[8] Or perhaps one should say that I *continue* the conversation initiated by the wrong.[9] As Murphy puts it (followed on this point by Hampton[10] and Hieronymi[11]):

> [I]njuries are also messages—symbolic communications. They are ways a wrongdoer has of saying to us, "I count but you do not," "I can use you for my purposes," or "I am here up high and you are there down below." Intentional wrongdoing insults us (…). (Murphy 1988, 25)

Responses can be a one-way process or a two-way process (Duff 2025, 93-94): it is a one-way process when no response is expected from the wrongdoer or when the intention is to terminate the conversation immediately (as in case 1.C when the victim kills the wrongdoer), and it is a two-way process when the response seeks a response (as in cases where we expect, or explicitly demand, an apology). By contrast, reactions can

[6] Narayan's interesting distinction between censure and denunciation (expressive theories) can be seen as a particular instantiation of the more general distinction I propose between responses and reactions. The distinctive characteristics she gives overlap largely with my own.

[7] It is also the word employed by Narayan in her justification for censure, by opposition with mere denunciation: "We are interested in *confronting* the perpetrator with the judgement that she is a responsible wrongdoer (…)" (Narayan 1992, 172).

[8] In general, my analysis has much in common with McKenna's "conversational theory of moral responsibility" (McKenna 2012).

[9] "Responding to an agent's action (…) is like engaging in an unfolding conversation with the agent whose act can be thought of as the initiation of a conversation" (McKenna 2012, 213).

[10] "By victimizing me, the wrongdoer has declared himself elevated with respect to me, acting as a superior who is permitted to use me for his purposes. A false moral claim has been made. Moral reality has been denied" (Hampton 1988, 125).

[11] "[A] past wrong against you (…) makes a claim. It says, in effect, that you can be treated in this way, and that such treatment is acceptable" (Hieronymi 2001, 546).

leave the wrongdoer completely out of the conversation ("as if her participation in any discourse about the wrongdoing (…) is pointless" says Narayan 1993, 172). In fact, reactions can be seen as a way of avoiding or bypassing a conversation that we do not want to have for various reasons, including "the strains of involvement" (Strawson 2008, 10).

It is worth noting that the confrontational aspect of responses does not mean that they are always retaliatory. The concept of response should not be confused with the concept of retaliation/reprisal. The noun 'retaliation' comes from the Latin verb *retaliare* (to pay back in kind), which is also found in the *lex talionis* (an eye for an eye, a tooth for a tooth): retaliation thus implies a form of paying back: I hurt you because you hurt me. In contrast, responses to a wrong can be gentle and pacifying: we can respond to a wrong with dialogue, without intending to harm the wrongdoer.

The confrontational nature of responses also explains why we may be reluctant to respond to wrongdoing in certain circumstances: we may not want to respond, for instance, to racial harassment either because we do not see the point of confronting the wrongdoer (who, we are sure, will never understand why what he did was wrong), or because we think confronting him would be dangerous (risk of escalation).

At this point, at least three apparent counter-examples might be put forward: conviction in absentia, transitional justice, and blaming/forgiving the dead. Let us take them in turn.

In 1940, in occupied France, the Pétain regime sentenced de Gaulle, then in exile in England, to death in absentia for treason. Does this show that (penal) responses need not be confrontational? Not really, because in the case of a conviction in absentia, the conviction is addressed to the culprit (and even beyond to the public at large, as I explain below): it is made public so that he can learn about it (even if his whereabouts are sometimes unknown). The Pétain regime knew that de Gaulle was in England and would be informed of this death sentence. It acted more or less like a person searching in the dark for someone hiding and threatening "If I find you, you will regret it".

Transitional justice also appears to be a counter-example. The Algerians, for example, have long demanded an apology from France for 132 years of colonization, for the massacres in Setif on May 8, 1945 that left thousands dead, for the use of torture during the Algerian war, etc. Demanding an apology seems to be a form of (non-violent) response to wrongdoing. But the problem is that many of the key perpetrators are deceased. To this we can answer that the Algerians are demanding an apology from the French state, which has not ceased to exist since the commission of these crimes and whose continuity has been ensured by the succession of different governments. In other words, the addressee of this demand is the French state or its representatives. It is up to them to acknowledge the wrongdoing that gave rise to this demand. On the other hand, if France were to be wiped off the map as a result of an invasion, Algeria would not be in a position to respond to the crimes committed by France and, above all, to demand an apology.

As for blaming/forgiving the dead, this can be seen as a response to the extent that we believe that they somehow hear us (the rationality of this belief is not our concern here). We could also simply say that "[t]hese are derivative cases to be accounted for in terms of how we would respond to the dead were they still alive" (McKenna 2012, 177).

Responses not only have an addressee, but they also convey a message. In other words, they are "communicative" (Macnamara 2015). When I merely react to a wrong, as the B series shows, I do not try to tell the wrongdoer something, to make him understand something. I mainly try to stop the wrong, to prevent it from happening again. What is said by a response—the message conveyed—depends, of course, on the nature of the response, although they probably all express at least some judgment of disapproval, such as "What you did was (what you are doing is) wrong". For instance, if you respond to a wrong (e.g. infidelity) with forgiveness, telling the wrongdoer "I forgive you", you probably mean that *although* what he did was wrong, you do not hold it against him.

Because responses are communicative, they can be misinterpreted or difficult to interpret. Consider the example of Jesus turning the other cheek in the Bible. This is a very baffling and unexpected response to a slap. What is it supposed to mean? Perhaps something like "I am against violence: you can slap me as much as you want, I will not respond with violence, I will not play your game", but other interpretations are possible, as theological controversies show.

Because responses involve an addressee, a dialogical dimension, and a communicative function, they are generally not directed at entities that lack understanding or *logos*, such as trees or flies (Macnamara 2015, 2, 17). Responses "make sense only on the assumption that the other can comprehend the message" (Watson 2019, 230). The more limited the other's level of understanding, the more limited the possibilities of response. That is the reason why we cannot respond to the aggressive behavior of a young child as we can to that of an adult.

By contrast, we can react to damage caused by entities that lack understanding. Although you will not respond to rat damage in your kitchen (by getting even with the rats or forgiving them), you may react by calling a pest control company; although you will not respond to rain damage in your ceiling (by blaming the rain), you may react by cleaning out your gutters.

We can also react to harm caused by people whose understanding is severely impaired. In the recent Sarah Halimi case in France, the man who killed this elderly woman was declared not guilty by reason of insanity and was consequently sent to a psychiatric hospital and subjected to "security measures" (*mesures de sûreté*) for the next twenty years. The state, or more precisely the criminal justice system, reacted to this crime (took therapeutic and security measures) but did not respond. That is why many people were dissatisfied with this decision: they wanted Kobili Traoré to be sentenced to life imprisonment; they wanted a strong and severe response, and they only got a proportionate and rational reaction.

One might object that, just as we sometimes get angry at an inanimate object, say a chair, after bumping into it, so too we sometimes respond to a "wrong" caused by nonhuman beings. Seneca gives the example of Cyrus the Great who tried to ford an Iranian-Iraqi river, but whose chariot was swept away by the current. Outraged, he had 180 channels dug, dividing the river into 360 streams (Seneca 2012, 80-81). However, in this example, Cyrus the Great punished the river as if it were a responsible agent that had shown insolence. He taught it a lesson as if it could understand the lesson:

> We tend to think that we have a right to expect 'respect' and cooperation from the inanimate objects that serve our ends, and in the moment we react as if they were bad people, since they clearly are not doing 'their job' for us. (Nussbaum 2016, 18-19)

Such conduct is, of course, childish, not to say pathetic, as Seneca points out. It shows how anger can lead to completely irrational responses.

Responses and reactions do not have the same success conditions.[12] What I mean here by "success" needs to be clarified. An act of revenge can succeed at being a revenge, at avenging for instance my dead brother (logical/conceptual success), but it can also succeed at accomplishing various aims: deter, make suffer, and so on (teleological/instrumental

[12] I would like to thank reviewer 2, who greatly helped me clarify this entire section.

success).[13] I am not here concerned with the conditions for teleological success, but only with the conditions for logical success: why some of our attempts at responding to a wrong fail to get beyond the attempt stage? The thesis I will defend is that for our response to succeed in the logical sense, it has to be identified as such, to be received as a response by the wrongdoer. The wrongdoer must *understand* that our response is a response *to the wrong we attribute to him*—he does not have to believe that he has committed a wrong, as long as he is aware that we think he has (this means that moral disagreement is compatible with responses). For instance, if someone takes revenge for a wrong by attacking the aggressor, but the aggressor has lost his memory due to a degenerative disease and does not remember the wrong he committed, the revenge cannot succeed, the response fails. Here is another example: if I demand an apology from a person, but that person asks me why, my response—the demand for an apology—has failed. In contrast, reactions to wrongdoing do not have to be identified as such by the wrongdoer in order to succeed as reactions; they operate independently of the wrongdoer's awareness and understanding.

A counterexample here might be drawn from the film *Old Boy*, where the main character, Dae-su, is kidnapped and wakes up in a sealed hotel room without knowing why. He is released 15 years later, but still ignores why he was kidnapped. He later understands that this torture was part of Woo-jin's revenge for some revelations he made in high school, which had disastrous consequences. But let us assume for the sake of the argument that he never learns why he was kidnapped. In that case, can we say that Woo-jin responded to Dae-su's wrong? From Dae-su's point of view, this would not be a response; or maybe he would speculate that it is a response for something he did, though he does not know what, or that he is the victim of a strange experiment without his consent. But from Woo-jin's point of view? My intuition tells me that his revenge would not be fully satisfying in this scenario: if Dae-su believes that his torture is not a response to his revelations, but just an experiment planned by the government or a secret society of psychologists, i.e. if he has *false beliefs* about the reasons why he is tortured, the revenge somehow fails in the logical sense—it is an attempted but unsuccessful revenge. This is probably why, in any revenge scenario, the avenger makes sure, or takes it for granted, that the offending party understands that he is getting even.

[13] Duff seems to rely on this distinction, though he does not make it explicit: "Communication seeks a particular kind of response from its addressee: if you do not hear or understand me, my communication has failed. It often also seeks not just understanding, but acceptance: I intend, or hope, that you will accept what I communicate, and my enterprise has been to that extent a failure if you are unpersuaded; but so long as you understand, I communicate successfully" (Duff 2025, 94). In my terminology, the understanding condition is a logical success condition: my attempted communication fails at being a communication if you do not understand me. However, the persuasion condition is a teleological success condition: communication did not achieve its objective if you are unpersuaded.

It is worth noting that Aristotle makes a very similar observation in his *Rhetoric*. He explains that we do not punish people when we know that they "will not perceive who is the cause of their suffering and that it is retribution for what [we] have suffered". He then uses the example of Odysseus and Polyphemus the Cyclops. After gouging out Polyphemus' eye and rejoining his ship, Odysseus reveals his true identity. This seems to be an act of hubris on the part of Odysseus, but Aristotle thinks that it is perfectly understandable "since [Odysseus] would not have been avenged if [Polyphemus the Cyclops] had not realized both from whom and why revenge came" (Aristotle 2007, 123, 1380b).

Cases in which the response is not identified as such should be distinguished from cases in which the response is identified as such but is completely ignored. For instance, the young child in 3.C may respond to the parents' scolding by smiling or even laughing, pick up a new book and start to tear it up. In 2.C the harasser may ignore the bystander's threats, act as if the bystander does not exist, and continue to harass his victim. One last example is what Duff calls the "defiant offender": "the offender that will not even listen to the moral message that his punishment seeks to communicate" (Duff 2001, 123). In all these cases, the response did not have the effect we wanted it to have (which was to stop the wrongdoing, or to engage the wrongdoer in a serious consideration of what he has done), even though a response proper took place—teleological failure, but logical success.

As we can see, responses to succeed in the logical and teleological sense depend largely on the wrongdoer. Which is not surprising: a conversation can succeed in both senses only if each speaker is able to discuss and is open to discussion. Which also means that responses are vulnerable and need to be backed up by reactions: in 2.C, faced with this failure, the only thing left to do is call the police; in 3.C, faced with this failure, the only thing left to do is move the books out of reach.

My account of responses, though is incorporates insights from expressivism (most notably, Feinberg's 1968 expressive account of punishment), should be distinguished from the latter; the fact that your reaction to a wrong expresses, say, your hatred does not suffice for it to be a response in the strict sense, since the expression of hatred may be devoid of any confrontational-communicative dimension and lack the identification component described above. As Duff—who also wishes to distinguish his communicative theory of punishment from expressive ones and to whom I owe this point[14]—puts it:

[14] For a similar point, see also Narayan (1993).

> I can express, and gratify, my hatred by harming the person I hate, and it might not matter to me whether he knows that I did this: I sabotage my hated enemy's car, and am gratified when he crashes, even if he believes the crash was accidental. Sometimes, however, it matters that the object of my expressive action knows about it: I want him to realize not simply that he is harmed, but that I have harmed him. (Duff 2025, 93)

Now that I have highlighted the main differences between responses and reactions, we can more confidently assess what qualifies as a response and what does not. For instance, in her list of responses other than punishment, Walker includes commemoration (Walker 2012, 10), yet, I would argue that commemoration fits more accurately within the category of reaction, that is within the B series. Commemoration bears all the distinctive marks of a reaction: it is opposed to indifference or inaction; it aims to prevent the repetition of the past; it is seen as a civic duty (like calling the police upon witnessing sexual harassment, as in case 2.B); we can commemorate natural disasters (Japan, for example, commemorates the Fukushima disaster) when we cannot respond to them (unless we personify nature as responsible and guilty); it does not need to be recognized as such by the perpetrators of the tragedy being commemorated; it does not imply a communicative confrontation with the author, but rather a communion between us: it is addressed above all to the younger generations.

Although reactions and responses are different, the boundary between them is not always clear-cut. In fact, the same behavior can be considered a reaction or a response depending on factors like context or the agent's mental state. For instance, cutting ties with someone can be understood in two different ways. As a response, it might be intended as a form of punishment, aiming to impose distress or teach a lesson, either as an end in itself or as a means to achieve some other outcome (it thus fits within the C series). In contrast, cutting ties might also be a reaction—an act of self-protection in the face of a toxic relationship, where the painful effects on the other person, though anticipated, are not the intended goal but rather an unfortunate byproduct (it thus fits within the B series). In this second case, the separation tends to be more abrupt and definitive. When cutting ties is intended as a punishment, there is at least some underlying communicative intent, which, paradoxically, maintains the ties by cutting them. This nuanced difference may help explain why so-called *pervers narcissiques* (a French term for manipulative narcissists) often seem to prefer being punished rather than shunned.

4. Applications

4.1 Responses and responsibility

Some authors have pointed out that the word “responsibility” originates from the Latin *respondere* which, before meaning to answer questions, meant to answer accusations. Thus, etymologically, responsibility means “answerability”:

> [A] person who is responsible for something may be required to answer questions (…). To say that a minister is responsible for the conduct of his department implies that he is obliged to answer questions about it if things go wrong (…). The original meaning of the word “answer” (…) was not that of answering questions, but that of answering or rebutting accusations or charges, which, if established, carried liability to punishment or blame or other adverse treatment (…). (Hart 2008, 265)

But responsibility can also be associated with “responses”, not in the sense of verbal answers to questions/accusations, but in the sense of actions motivated by a wrong and addressed to the wrongdoer. In other words, it can also be defined as “responseworthiness”.[15] According to this definition, the word “responsible” parallels the word “desirable” (worthy of being desired). To be responsible for a wrong is “to be eligible to a range of (…) responses with a built-in confrontational element” (Shoemaker 2015, 87).

Thus, an important difference between responsibility as “answerability” and responsibility as “responseworthiness” is that in the first case verbal answers go from the agent to others (the person who is responsible is required to answer questions, accusations), while in the second case behavioral responses go from others to the agent (the person who is responsible is liable to responses such as blame, “moral protest”, punishment, forgiveness, etc.).

Although distinct, answerability and responseworthiness, at least in the criminal justice context,[16] are related as follows: “a person who fails to rebut a charge is liable to punishment or blame” or some other response

[15] “[I]n the broad sense, moral responsibility for an action is a matter of *deserving a moral response* on the basis of the action” (Copp 1997, 452).

[16] I say “at least in the criminal justice context”, because outside of that context we may respond to what we perceive as a wrong without giving the addressee an opportunity to defend himself. When McKenna says that punishment is “a response that might be regarded as fitting *after* a morally responsible agent has had an opportunity to offer an account of her conduct” (2012, 91), he is making a normative claim about when punishment should be inflicted.

"for what he has done, and a person who is liable to punishment or blame" or some other response "has had a charge to rebut and failed to rebut it" (Hart 2008, 265). In other words, the answerable person becomes a responseworthy person if, when accused of a wrong, he is unable to defend himself with a compelling justification or excuse. And if he could not defend himself with a compelling justification or excuse, but acknowledged the wrong done and apologized, he is worthy of a merciful response.[17]

One final remark: the fact that a person is responsible for a wrong in the sense of responseworthiness does not mean that we should respond to his wrong, nor that failing to respond and merely reacting or resenting him would necessarily be blameworthy or inappropriate. It rather means, to borrow Shoemaker's term cited earlier, that the person is eligible to responses—that they are grounds for punishing, blaming, demand an apology, etc. But of course, people do not always get what they are eligible to—grounds for responding can be defeated or overridden by other, competing considerations.

4.2 Penal responses

Penal responses constitute a subset of responses and this raises the question: what is specific about them? It might seem that penal responses, as their name suggests, are defined by their punitive nature, implying a narrower scope than responses in general. However, this assumption is incorrect. The criminal justice system sometimes employs non-punitive responses, such as issuing a legal reminder, granting an absolute discharge (*dispense de peine* in French law), or mandating participation in rehabilitative programs, such as those for substance abuse in drug courts.

In fact, penal responses are characterized not by a single feature, but by several. First, they are structured rather than diffuse; they are administered by an organ of the state—namely, the judiciary (Durkheim 2013, 55). As a result, penal responses involve "coercion that is unilateral only": the guilty "is subjected to an external duty to which he, for his own part, may offer no resistance", and this may offend his "feeling of honor" or suspend "his dignity as a citizen" (Kant 1991, 168, note). A convicted person, of course, may appeal, but if the conviction is upheld, he must submit to it, he cannot resist it. Besides, the target of penal responses (what they respond to) is narrower than the target of responses in general: they respond only to

[17] The question of the conditions for responseworthiness, like the question of the conditions for blameworthiness, is a complex one that I cannot resolve in this paper. However, it seems plausible that there are both general conditions for being responseworthy and more specific conditions for being worthy of this or that type of response.

criminal wrongdoing (offenses), not to wrongdoing in general. Finally, the addressee of penal responses is not only the offender, but also the public at large.[18] Certainly, some cases of private revenge involve a kind of extended communication: revenge is addressed not only to the aggressor, but also to those around him who are warned "Do not imitate him". It can also be found in classrooms where the teacher's responses to disciplinary offenses are addressed not only to the offender, but also to the entire class. However, penal responses go further: for instance, when Harvey Weinstein was sentenced to 23 years, society as a whole was reminded not only of the offense's gravity but also of the legal and moral condemnation of sexual violence.

Interestingly, some proponents of retributivism or "just deserts" paradoxically defend one penal response, legal punishment, as more humane than therapy or reform, which they view as degrading, i.e. bestializing, infantilizing, or demonizing (Waldron 2010, 282-83). Consider, for instance, the following excerpt from "The Humanitarian Theory of Punishment" by Lewis:

> To be "cured" against one's will and cured of states which we may not regard as disease is to be put on a level with those who have not yet reached the age of reason or those who never will; to be classed with infants, imbeciles, and domestic animals. But to be punished, however severely, because we have deserved it, because we "ought to have known better", is to be treated as a human person made in God's image. (Lewis 1987, 151)

Lewis holds that punishment treats the punished person "as a human person" probably because it is a response: no matter how serious the crime, punishment is *addressed* to the punished person as a creature capable of both speech and understanding ("made in God's image" and having "reached the age of reason") and sends the message that he "ought to have known better". In contrast, imposed cure has no addressee, but a recipient: it is a reaction that could also be appropriate for non-human creatures deprived of reason. From the point of view of its intentions, imposed cure seems to be more humane than punishment, since it does not seek to cause pain; but Lewis argues that, ontologically speaking, from the point of view of its nature, punishment is more humane than imposed cure.

[18] Like Duff (2025, 95) we can make a distinction between the primary addressee and the secondary addressee of a response.

However, there is a middle way between Lewis' retributivism and the Clockwork Orange scenario he opposes (in which crime is seen as a disease and punishment is replaced by an imposed cure). One might maintain that responding to wrongdoing is indeed essential if we want to treat the wrongdoer as a person, if we do not want "to give up on him as a moral agent" (Duff 2001, 123), while arguing that such responses need not be punitive in nature. It is for instance the view defended by Narayan in her censure theory: she argues that when we fail to *censure* the wrongdoer and merely react to his wrong (for instance by denouncing it), we treat "the actor as one would treat a natural source of trouble—viz, as an entity not to be directly addressed" (Narayan 1993, 172).

One final remark regarding penal responses. What our analysis reveals is that most definitions of punishment are deficient, as they fail to include the identification component intrinsic to any type of response in the strict sense. This component is not included in the Flew-Benn-Hart definition, i.e. pain intentionally inflicted by the state on an offender for his offense (Hart 2008, 4-5); it is not included in Boonin's definition of legal punishment as authorized intentional reprobative retributive harm (Boonin 2008, 26). Yet this element is necessary if we are to avoid puzzling cases being classified as punishment. Let us consider, for instance, the cases of animal trials in medieval Europe and trials of inanimate objects in Ancient Greece. Will we say that authorized intentional reprobative retributive harm inflicted upon animals and inanimate objects count as punishment for, say, murder? If we are reluctant to say that they are or were punished for murder, we now have a clear explanation: we tend to understand punishment as a response to wrongdoing; now, animals and even more so inanimate objects are unable to perceive harm inflicted upon them after trial as a response to the wrong we attribute to them.

5. Strawsonian reactive attitudes and practices

An objection might be raised that I am merely reiterating Strawson's distinction between the objective attitude and the participant attitude; between objective attitudes and reactive attitudes. It could be argued that I am expressing Strawson's view in alternative terms. However, I believe this interpretation is mistaken, and I will clarify why.

First, you may choose to react to a wrong rather than respond to it without adopting an objective stance toward the wrongdoer. Just because you merely react to a crime—for instance by calling the police, as in case 2.B, or by commemorating it—does not mean that you see the perpetrator (or the perpetrators) with an objective eye, "as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of" (Strawson 2008, 9).

Second, Strawson contends that "reactive attitudes are essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in their attitudes and actions" (Strawson 2008, 10-11). Yet this observation does not wholly apply to responses in either interpersonal or criminal justice contexts. You may blame someone who broke a vase out of negligence by telling him that he should have been more thoughtful, though his action shows no ill will, just a lack of reasonable care (Hart 2008, 136). Similarly, strict liability, however unfair it may be, shows that we may be punished with small fines for many actions (e.g. selling expired products through no fault of our own) without being blameworthy. However, the weaker point—the commonplace—that our responses are *influenced* by our beliefs about "the quality of others' will towards us" (Strawson 2008, 15) is obviously true.

Third, Strawson claims that "reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard (…) towards ourselves" (Strawson 2008, 15). While this is accurate in some instances, it cannot be generalized to all. When you respond negatively to a personal injury or an action that has hurt your feelings, the underlying demand for regard is easy to see. However, when the criminal justice system responds to the crime of tax evasion with fines or imprisonment, this demand is less easy to perceive. Similarly, some student disciplinary offenses (such as falsifying documents) may require certain responses that are not based on a demand of regard towards anyone—rather, they are based on an expectation of compliance with the rules that has not been met.

Fourth, Strawson (and most of the post-Strawsonian literature) is mainly interested in our human "moral sentiments" (Strawson 2008, 26), whereas I am, as I have emphasized, primarily interested in our behaviors—in what we *do* in reaction or in response to wrongdoing, not in what we *feel* in reaction to wrongdoing. By no means do I wish to deny that what we do and what we feel are closely related. Imagine that you have been wronged and are considering how to respond. You decide to go and see the person who wronged you to explain that you are upset with him, or to show him that you are angry. Even if emotions play a central role in the response you choose, it is still a response in the strict sense (fitting within the C series), in that you are not just brooding over your anger or digesting your resentment: you are confronting the person who wronged you to let him know how you feel.

Fifth, the concept of responses I propose aligns more closely with what Strawson calls the (reactive) “practices” of blame, moral condemnation, and punishment, which he sets aside at the beginning of “Freedom and Resentment” only to reconsider them at the end. One of Strawson’s main points is that we should not lose sight of the fact that, whether we like it or not, these practices are in part expressions of our moral sentiments; they are “not merely devices we calculatingly employ for regulative purposes” (Strawson 2008, 27). This insight applies perfectly to responses to wrongdoing. When we discuss the rationale for a given response (or a type of response), we may put forth its beneficial consequences, such as reducing recidivism. However, this does not mean that this response (or type of response) is a tool designed to produce these beneficial consequences. In fact, it will rarely be perceived as such by the respondent, and even less so by the addressee. What we see in a punishment inflicted on us is primarily the emotion it expresses (e.g. resentment) and the message it conveys (e.g. you “ought to have known better” as C. S. Lewis would say). Even if it is inflicted for our own good, that is not what it immediately means to us.

Sixth, Strawson’s argument against the feasibility of adopting a thoroughly objective attitude applies equally to reactions. It seems, “for us as we are, practically inconceivable” (Strawson 2008, 12) to abandon responses in favor of reactions; to react to harm caused by others exclusively in the same way as we react to damage caused by entities deprived of *logos*; by taking only measures that are not addressed to anyone, that have no dialogical dimension, that do not involve looking each other in the eye. Besides, such a world—which is not ours and never will be—would greatly impoverish human life (Strawson 2008, 14). Most would find this world unappealing, as it would lack the richness of a truly human community.

6. An assessment of response retributivism

We are now ready to answer the question that constitutes the title of this article. A response to moral or criminal wrongdoing is a form of action motivated by the wrong. Its defining characteristic is that it is addressed to the wrongdoer, has a confrontational-communicative dimension, and needs to be identified by the wrongdoer as a response to his wrong. Thus, it is perfectly possible to react to a wrong without responding to it. And it seems to me that one of the many meanings of the ambiguous word “responsible” refers precisely to responses in this strict sense: responsible agents are those who are “responseworthy”.

These reflections on the nature of responses to wrongdoing may shed light on the question of what constitutes an *appropriate* response to wrongdoing. For instance, my analysis suggests that letting go or feeling resentment cannot be classified as appropriate responses to wrongdoing; they are not responses to wrongdoing at all, though one could argue that letting go is sometimes the right thing to do, and that resentment may be an appropriate emotion. The question of the appropriate response to wrongdoing, if we use the term "response" in a vague and indeterminate way, is extremely difficult to address. It would probably be easier to address if we made a more rigorous distinction, as I propose, between: first, the question of the appropriate emotion in response to wrongdoing—*is it appropriate to feel pity for the perpetrator of a mass murder in a school?* Second, the question of the appropriate reaction—*is installing metal detectors at schools an appropriate reaction?* Third, the question of the appropriate response (in the strict sense)—*is it appropriate to sentence the perpetrator of a mass murder to an educational measure?* Furthermore, such a distinction would spare us the unproductive search for *the* appropriate response (in the vague and indeterminate sense); it would compel us to acknowledge that, when faced with wrongdoing, several things can be appropriate at the same time: a given emotion, a given security measure, and a given confrontational-communicative act.[19]

This distinction may also help us to evaluate the merits and limitations of response retributivism. With regard to individual non-state actors, the idea that there is a general duty to respond to wrongdoing seems implausible to me. Responses are sometimes too dangerous because of their confrontational nature; in other cases, they are too painful, especially for victims who no longer wish any contact with their aggressor; in other cases, there is no doubt that they will be ineffective; in still other cases, we may lack the standing or authority required to respond to wrongdoing: "X's mother may have the standing to censure her for stealing cookies out of her cookie-jar, but X's busybody neighbour does not" (Narayan 1993, 168). The Aristotelian notion that refusing to respond (especially to take revenge) is a sign of servility is simply a reflection of a culture of honor that is no longer ours. Still regarding individual non-state actors, I am not even sure that there is a general duty to react to wrongdoing: sometimes, the best thing to do may be to do nothing. As Brunning and Milam note:

> We let go for particular reasons, in response to particular situations of moral conflict, and often with particular aims. And these are often good reasons to do so. When it accomplishes

[19] I would like to thank reviewer 2 for bringing this point to my attention.

> these aims, letting go can be good (...): therapeutic, liberating, beneficent, and even virtuous. (Brunning and Milam 2022, 15)

That said, even if there is no general duty, I do not wish to deny that in certain specific contexts or situations, individual non-state actors may have a duty to react, or even to respond. The clearest case arises when they bear a "role-responsibility" (Hart 2008). For example, if you are responsible for the safety of a group of children, you have at least a duty to react when one of them is wronged. Likewise, as a parent, it is part of your role to respond to serious wrongdoing committed by your child, provided they are sufficiently mature to understand what you are attempting to communicate.

Now, what about the state? Four cases should be distinguished:

- Response with reaction: The criminal justice system punishes the offender, supplements the sentence with security measures and a treatment order, while the government launches a broad initiative to eliminate or significantly reduce this type of crime.
- Response without reaction: The wrong is minor, and the state considers that a warning is the most appropriate response, requiring no further action.
- Reaction without response: The perpetrator of the murder suffers from serious mental disorders and is unaware of what he has done, rendering him "unfit to stand trial". In such a case, responding to the wrong he has done would be meaningless (Duff 2025, 94); the person is instead committed to a psychiatric facility.
- Absence of reaction and response: According to the principle of prosecutorial discretion (the "opportunity principle" in French criminal law), a prosecutor may decide not to prosecute an apparent violation of criminal law, since he considers that prosecution would likely do more evil than good (see Dempsey 2009).

As we can see, the state does not have a general duty to react or respond to (legal) wrongdoing, surprising as it may seem. One might object that in *normal* cases, the state has a *pro tanto* duty to respond to (legal) wrongdoing. But such a thesis is too vague to be normatively interesting; it raises more questions than it resolves: What constitutes a normal case? Which response is to be chosen? Which circumstances can defeat or override this *pro tanto* duty? Thus, the truly difficult questions for state actors are rather the *when* question and the *which* question: When to drop charges? When to complement punishment with safety measures, and which ones? When to use an alternative response to punishment, and which one? However, these questions are beyond the scope of this article.

Acknowledgments

This first version of this article was written as part of the REACT project, funded by the Agence Nationale de la Recherche (ANR-20-CE28-0012). I presented it at the German-French-Italian Workshop "Themes on Moral Psychology" at LMU Munich in 2023. I am grateful to the participants, and in particular to Jörg Löschke, for their precious comments. Finally, I would like to thank the two anonymous reviewers for their helpful suggestions, which contributed to improving the article.

REFERENCES

Anscombe, G. E. M. 2000. *Intention*. Cambridge: Harvard University Press.

Aristotle. 2007. *On Rhetoric: A Theory of Civic Discourse*. Translated by George A. Kennedy. New York: Oxford University Press.

Boonin, David. 2008. *The Problem of Punishment*. Cambridge: Cambridge University Press.

Brunning, Luke, and Per-Erik Milam. 2022. "Letting Go of Blame." *Philosophy and Phenomenological Research* 106: 720-740. https://doi.org/10.1111/phpr.12899.

Copp, David. 1997. "Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility." *Noûs* 31 (4): 441-56. https://doi.org/10.1111/0029-4624.00055

Dahan Katz, Leora. 2020. "Response Retributivism: Defending the Duty to Punish." *Law and Philosophy* 40 (6): 585–615. https://doi.org/10.1007/s10982-020-09386-3

Duff, Antony. 2025. "Communicative Theory." In *The Oxford Handbook of the Philosophy of Punishment*, edited by Jesper Ryberg, 90-112. Oxford: Oxford University Press.

Duff, Antony. 2001. *Punishment, Communication, and Community*. Oxford: Oxford University Press.

Durkheim, Émile. 2013. *The Division of Labour in Society*. Translated by W. D. Halls. Basingstoke: Palgrave Macmillan.

Feinberg, Joel. 1965. "The Expressive Function of Punishment." *The Monist* 49 (3): 397–423. https://doi.org/10.5840/monist196549326

Hampton, Jean. 1988. "The Retributive Idea." In *Forgiveness and Mercy*, by Jeffrie Murphy and Jean Hampton, 111-161. Cambridge: Cambridge University Press.

Hart, H. L. A. 2008. *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford: Oxford University Press.

Hieronymi, Pamela. 2001. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62 (3): 529-55. https://doi.org/10.1111/j.1933-1592.2001.tb00073.x.

Holmgren, Margaret Reed. 2012. *Forgiveness and Retribution: Responding to Wrongdoing*. New York: Cambridge University Press.

Kant, Immanuel. 1991. *The Metaphysics of Morals*. Cambridge: Cambridge University Press.

Lewis, C.S. 1987. "The Humanitarian Theory of Punishment." *Issues in Religion and Psychotherapy* 13 (1): 147-53.

Macnamara, Coleen. 2015. "Reactive Attitudes as Communicative Entities." *Philosophy and Phenomenological Research* 90 (3): 546–69. https://doi.org/10.1111/phpr.12075.

McKenna, Michael. 2012. *Conversation and Responsibility*. New York: Oxford University Press.

Metz, Thaddeus. 2008. "The Nature of Reactive Practices: Exploring Strawson's Expressivism." *South African Journal of Philosophy* 27 (3): 227-41. http://dx.doi.org/10.4314/sajpem.v27i3.31514

Murphy, Jeffrie. 1988. "Forgiveness and Resentment." In *Forgiveness and Mercy*, by Jeffrie Murphy and Jean Hampton, 14-34. Cambridge: Cambridge University Press.

Narayan, Uma. 1993. "Appropriate Responses and Preventive Benefits: Justifying Censure and Hard Treatment in Legal Punishment." *Oxford Journal of Legal Studies* 13 (2): 166-82. https://doi.org/10.1093/ojls/13.2.166

Nussbaum, Martha C. 2016. *Anger and Forgiveness: Resentment, Generosity, Justice*. New York: Oxford University Press.

Seneca. 2012. *Anger, Mercy, Revenge*. Translated by Robert A. Kaster and Martha C. Nussbaum. Chicago: University of Chicago Press.

Shoemaker, David. 2015. *Responsibility from the Margins*. Oxford: Oxford University Press.

Strawson, Peter F. 2008. *Freedom and Resentment and Other Essays*. Abingdon: Routledge.

Waldron, Jeremy. 2010. "Inhuman and Degrading Treatment: The Words Themselves." *Canadian Journal of Law & Jurisprudence* 23 (2): 269–86. https://doi.org/10.1017/S0841820900004938.

Walker, Margaret Urban. 2012. *Moral Repair: Reconstructing Moral Relations after Wrongdoing*. Cambridge: Cambridge University Press.

Watson, Gary. 2019. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Perspectives on Moral Responsibility*, edited by John Martin Fischer and Mark Ravizza, 119-48. Ithaca: Cornell University Press.

ABSTRACTS (SAŽECI)

Précis of Madness: A Philosophical Exploration

Justin Garson
Hunter College and The Graduate Center, City University of New York, USA

ABSTRACT

The following is a short synopsis of the book Madness: A Philosophical Exploration. It provides an overview of the book's core distinction between madness-as-dysfunction and madness-as-strategy, and enumerates four benefits of relying on this conceptual framework: for history, philosophy, Mad Pride, and treatment.

Keywords: psychiatry; mental disorder; madness-as-dysfunction; madness-as-strategy.

Précis knjige Madness: A Philosophical Exploration

Justin Garson
Hunter College and The Graduate Center, City University of New York, SAD

SAŽETAK

Slijedi kratak sinopsis knjige Madness: A Philosophical Exploration. Daje se pregled osnovne razlike koja se koristi u knjizi a odnosi se na ludilo kao disfunkciju i ludilo kao strategija, te se navode četiri benefita korištenja ovog pojmovnog okvira: za povijest, filozofiju, Mad Pride i liječenje.

Ključne riječi: psihijatrija; mentalni poremećaj; ludilo-kao-disfunkcija; ludilo-kao-strategija.

Madness by Design: A Genealogy of an "Anti-Tradition"

Muhammad Ali Khalidi
City University of New York, USA

ABSTRACT

Psychiatric conditions are commonly regarded as mental disorders or dysfunctions of the mind. Yet there is a wealth of historical theorizing about the mind that conceives of these conditions as, in some sense, a matter of design rather than dysfunction. This intellectual legacy is the topic of Justin Garson's penetrating study, Madness: A Philosophical Exploration (2022). In this paper, I interpret Garson's book as a genealogy (in the Foucauldian sense) of the "anti-tradition" that he labels "madness-as-design". I argue that viewing the intellectual legacy that Garson analyzes through this genealogical lens has two benefits. First, it encourages us to identify other instances of madness-as-design (or madness-by-design), particularly those with an overtly political dimension, such as psychiatric conditions in a colonial context. Second, it should lead us to question the category of madness itself, which turns out to be radically disjointed, particularly since it cannot be unified under the rubric of disorder or dysfunction.

Keywords: psychiatry; mental disorder; dysfunction; genealogy; colonialism.

Dizajnirano ludilo: genealogija „anti-tradicije”

Muhammad Ali Khalidi
City University of New York, SAD

SAŽETAK

Psihijatrijska stanja obično se smatraju mentalnim poremećajima ili disfunkcijama uma. Ipak, postoji bogatstvo povijesnog teoretiziranja o umu koje ove uvjete koncipira, u određenom smislu, stvarima dizajna, a ne disfunkcije. Ovo intelektualno naslijeđe tema je prodorne studije Justina Garsona, Madness: A Philosophical Exploration (2022). U ovom radu interpretiram Garsonovu knjigu kao genealogiju (u Foucauldijevom smislu) „anti-tradicije“ koju naziva „ludilo kao dizajn“. Tvrdim da gledanje na intelektualno naslijeđe koje Garson analizira kroz ovu genealogijsku prizmu ima dvije koristi. Prvo, potiče nas da prepoznamo druge primjere ludila kao dizajna (ili dizajniranog ludila), osobito one s otvoreno političkom dimenzijom, poput psihijatrijskih stanja u kolonijalnom kontekstu. Drugo, to bi nas trebalo navesti na preispitivanje same kategorije ludila, koja se pokazuje radikalno razjedinjenom, osobito jer se ne može ujediniti pod pojmom poremećaja ili disfunkcije.

Ključne riječi: psihijatrija; mentalni poremećaj; disfunkcija; genealogija; kolonijalizam.

Strategy, Pyrrhonian Scepticism and the Allure of Madness

Sofia Jeppsson
Umeå University, Sweden

Paul Lodge
University of Oxford, United Kingdom

ABSTRACT

Justin Garson introduces the distinction between two views on Madness we encounter again and again throughout history: Madness as dysfunction, and Madness as strategy. On the latter view, Madness serves some purpose for the person experiencing it, even if it's simultaneously harmful. The strategy view makes intelligible why Madness often holds a certain allure—even when it's prima facie terrifying. Moreover, if Madness is a strategy in Garson's metaphorical sense—if it serves a purpose—it makes sense to use consciously chosen strategies for living with Madness that don't necessarily aim to annihilate or repress it as far as possible. In this paper, we use our own respective stories as case studies. We have both struggled to resist the allure of Madness, and both ended up embracing a kind of Pyrrhonian scepticism about reality instead of clinging to sane reality.

Keywords: madness; Pyrrhonian scepticism; mania; psychosis; psychiatry.

Strategija, pironovski skepticizam i privlačnost ludila

Sofia Jeppsson
Umeå University, Švedska

Paul Lodge
University of Oxford, Ujedinjeno Kraljevstvo

SAŽETAK

Justin Garson uvodi razliku između dva stajališta o ludilu koja se iznova pojavljuju kroz povijest: ludilo kao disfunkcija i ludilo kao strategija. Prema drugom stajalištu, ludilo ima određenu svrhu za osobu koja ga doživljava, čak i ako istovremeno ima štetne posljedice. Strateška perspektiva objašnjava zašto ludilo često ima određenu privlačnost—čak i kada je na prvi pogled zastrašujuće. Štoviše, ako je ludilo strategija u Garsonovom metaforičkom smislu—ako ima svrhu—tada ima smisla koristiti svjesno odabrane strategije za življenje s ludilom, koje ne moraju nužno težiti njegovom uništenju ili što je moguće većem potiskivanju. U ovom radu koristimo vlastite priče kao studije slučaja. Oboje smo se borili protiv privlačnosti ludila, te smo na kraju prihvatili određeni oblik pironovskog skepticizma u odnosu na stvarnost, umjesto da se držimo „normalne“ stvarnosti.

Ključne riječi: ludilo; pironovski skepticizam; manija; psihoza; psihijatrija.

Into the Deep End: From Madness-as-Strategy to Madness-as-Right

Miguel Núñez de Prado-Gordillo
University of Rijeka, Croatia

ABSTRACT

A central notion in Mad Pride activism is that “madness is a natural reaction” (Curtis et al. 2000, 22). In Madness: A Philosophical Exploration (2022), Justin Garson provides a compelling exploration and defence of this idea through the book’s central concept: madness-as-strategy, i.e., the view of madness as “a well- oiled machine, one in which all of the components work exactly as they ought” (1). This contrasts with the dominant view in 20th- and 21st-century psychiatry, madness-as-dysfunction, which understands madness as a failure of function. The paper provides a critical analysis of the notion of madness-as-strategy as a political tool, pointing out its main virtues and limitations in terms of Garson’s overarching political project: to carve out the conceptual landscape of madness in ways that pay tribute to mad people’s own perspectives. The analysis draws on two central commitments of contemporary neurodiversity theory: a) its relational-ecological model of cognitive (dis)ability; and b) its non-essentialist, sociopolitical critique of the “normalcy paradigm”. I argue that these two insights contribute to both expand the applicability of madness- as-strategy and highlight its limitations as a tool for the political struggles of mad, cognitively divergent, and mentally ill or disabled people. The paper concludes by outlining a way to move beyond both madness-as-dysfunction and madness-as-strategy, toward what I call madness-as-right.

Keywords: philosophy of psychiatry; conceptual explication; mad studies; neurodiversity paradigm; madness-as-dysfunction.

U duboku vodu: od ludila kao strategije do ludila kao prava

Miguel Núñez de Prado-Gordillo
Sveučilište u Rijeci, Hrvatska

SAŽETAK

Središnji pojam u Mad Pride aktivizmu jest da je „ludilo prirodna reakcija“ (Curtis i sur. 2000, 22). U knjizi Madness: a philosophical exploration (2022), Justin Garson pruža uvjerljivo istraživanje i obranu ove ideje kroz glavni pojam knjige: ludilo kao strategija, tj. perspektivu na ludilo kao „dobro podmazan stroj, u kojem svi njegovi dijelovi rade upravo onako kako bi trebali“ (1). To je u suprotnosti s dominantnim gledištem u psihijatriji 20. i 21. stoljeća, koje se referira na ludilo kao disfunkciju. Rad pruža kritičku analizu pojma ludilo-kao-strategija kao političkog alata, ukazujući na njegove glavne vrline i ograničenja u okviru Garsonovog općeg političkog projekta: oblikovati pojmovni prostor ludila na načine koji odaju počast perspektivama ljudi s ludilom. Analiza se oslanja na dva središnja temelja suvremene teorije neurorazličitosti: a) njezin relacijski-ekološki model kognitivnih invaliditeta; i b) njezinu ne-esencijalističku, sociopolitičku kritiku „paradigme normalnosti“. Tvrdim da ova dva uvida doprinose proširenju primjenjivosti ludila-kao-strategije i ističu njezina ograničenja kao alata za političke borbe ljudi s iskustvom ludila, kognitivno divergentnih, te mentalno oboljelih ili osoba s invaliditetom. Rad zaključujem iznošenjem načina na koji se može prevazići ludilo-kao-disfunkciju i ludilo-kao-strategiju, prema onome što nazivam ludilo-kao-pravo.

Ključne riječi: filozofija psihijatrije; pojmovna eksplikacija; studije ludila; paradigma neurorazličitosti; ludilo-kao-disfunkcija.

Reconceptualizing Delusion: Strategy, Dysfunction, and Epistemic Injustice in Psychiatry

Eleanor Palafox-Harris
University of Birmingham, United Kingdom

Ema Sullivan Bissett
University of Birmingham, United Kingdom

ABSTRACT

In his bold and illuminating book Madness: A Philosophical Exploration, Justin Garson makes a case for thinking about madness as strategy, rather than as dysfunction. The reader is invited to take away a better appreciation of the historical provenance of madness as strategy, that is, this is not a new idea, destined for the fringes or of interest only to those of a more radical bent. It is rather an idea which has firm roots in the history of psychiatry. Garson's lens is wide, he is advocating a strategy over dysfunction approach for, at least, anxiety, depression, schizophrenia (and its spectrum disorders), and delusion. In this exploratory paper, we focus on delusion. We discuss what a madness-as-strategy approach might say about delusion, and how that fits with the idea that such beliefs are evolutionarily adaptive. We turn then to explore the implications of this reconceptualization of delusion for epistemic injustice in psychiatry. Our discussions will support the idea that much of the theoretical action lies not in the distinction between dysfunction and strategy, but rather in the distinction between everyday and abnormal dysfunction.

Keywords: delusion; strategy; dysfunction; doxastic dysfunction; abnormality; epistemic injustice.

Rekonceptualizacija deluzije: strategija, disfunkcija i epistemička nepravda u psihijatriji

Eleanor Palafox-Harris
University of Birmingham, Ujedinjeno Kraljevstvo

Ema Sullivan Bissett
University of Birmingham, Ujedinjeno Kraljevstvo

SAŽETAK

U svojoj hrabroj i prosvjetljujućoj knjizi *Madness: a philosophical* *exploration*, Justin Garson zagovara shvaćanje ludila kao strategije, a ne disfunkcije. Čitatelj je pozvan da stekne bolje razumijevanje povijesnog podrijetla ludila kao strategije, što znači da nije riječ o novoj ideji koja je namijenjena rubnim područjima ili je zanimljiva samo onima s radikalnijim sklonostima. Naprotiv, riječ je o ideji koja ima čvrste korijene u povijesti psihijatrije. Garsonov doseg je širok; on zagovara pristup strategije umjesto disfunkcije, barem u slučaju anksioznosti, depresije, shizofrenije (i njezinih spektralnih poremećaja) te deluzije. U ovom radu fokusiramo se na deluzije. Raspravljamo o tome što bi pristup ludilu kao strategiji mogao reći o deluzijama i kako to odgovara ideji da su takva vjerovanja evolucijski adaptivna. Zatim istražujemo implikacije ove rekonceptualizacije deluzija u kontekstu epistemičke nepravde u psihijatriji. Naša rasprava podupire ideju da se puno teorijske akcije zapravo nalazi, ne u razlici između disfunkcije i strategije, već u razlici između svakodnevne i abnormalne disfunkcije.

Ključne riječi: deluzije; strategija; disfunkcija; doksastička disfunkcija; abnormalnost; epistemička nepravda.

Madness Revisited: Replies to Contributors

Justin Garson
Hunter College and The Graduate Center, City University of New York, USA

ABSTRACT

The following provides the author's responses to the four commentaries on Madness: A Philosophical Exploration, written by Muhammad Ali Khalidi, Eleanor Palafox-Harris and Ema Sullivan-Bissett, Miguel Núñez de Prado Gordillo, and Sofia Jeppsson and Paul Lodge.

Keywords: mental disorder; natural kind; madness-as-dysfunction; madness-as-strategy; psychosis; delusion; genealogy**.**

Ponovno o ludilo: odgovori autorima

Justin Garson
Hunter College and The Graduate Center, City University of New York, SAD

SAŽETAK

Ovaj rad sadrži odgovore na četiri komentara na knjigu Madness: A Philosophical Exploration, koje su napisali Muhammad Ali Khalidi, Eleanor Palafox-Harris i Ema Sullivan-Bissett, Miguel Núñez de Prado Gordillo, te Sofia Jeppsson i Paul Lodge.

Ključne riječi: mentalni poremećaj; prirodna vrsta; ludilo-kao-disfunkcija; ludilo-kao-strategija; psihotičnost; deluzije; genealogija.

Habits and Dispositions in Frank Ramsey's Philosophy

Alice Morelli
Ca' Foscari University, Venice, Italy

ABSTRACT

This paper examines Ramsey's use of the concepts of habit and disposition, challenging the common interpretation that he employs them interchangeably in his theory of belief. This interpretative trend reflects a broader tendency to equate habit and disposition, based on the assumption that a habit is an acquired disposition to act. However, the precise relationship between these concepts often remains underexplored and it is not clear whether habits are merely a subset of dispositions or if they are conceptually distinct. Using Ramsey's writings as a case study, this paper argues that their relationship is more nuanced than a reductive equivalence suggests. I advance a twofold thesis: first, I argue that Ramsey's use of the notions of habit and disposition is more complex than typically assumed, as he employs them in distinct philosophical contexts and conceptualizes them in different ways. Second, I distinguish between a logical-grammatical kind of dispositionalism and a metaphysical one to argue that the notion of habit is dispositional but habits are not metaphysically equivalent to dispositions. Ramsey conceptualizes habits as methods, rules, procedures of thought, whereas dispositions are understood as tendencies or inclinations engendered and shaped by habits.

Keywords: Frank Ramsey; habit; disposition; pragmatism; normativity.

Navike i dispozicije u filozofiji Franka Ramseyja

Alice Morelli
Ca' Foscari University, Venecija, Italija

SAŽETAK

Ovaj rad analizira Ramseyjevu upotrebu pojmova navike i dispozicije, dovodeći u pitanje rašireno tumačenje prema kojem ih on koristi kao sinonime u svojoj teoriji vjerovanja. Taj interpretativni trend odražava širu tendenciju poistovjećivanja navike i dispozicije, temeljenu na pretpostavci da je navika stečena dispozicija za djelovanje. Međutim, precizan odnos između tih pojmova često ostaje nedovoljno razrađen, a nije jasno jesu li navike samo podskup dispozicija ili su pojmovno različite. Korištenjem Ramseyjevih tekstova kao studije slučaja, ovaj rad tvrdi da je njihov odnos nijansiraniji nego što to sugerira reduktivno poistovjećivanje. Izlažem dvostruku tezu: prvo, tvrdim da je Ramseyjeva upotreba pojmova navike i dispozicije složenija nego što se to obično pretpostavlja, jer ih koristi u različitim filozofskim kontekstima i pojmovno ih razlikuje. Drugo, razlikujem logičko-gramatički oblik dispozicionalizma od metafizičkog, kako bih argumentirala da je pojam navike dispozicionalan, ali da navike nisu metafizički ekvivalentne dispozicijama. Ramsey konceptualizira navike kao metode, pravila i misaone procedure, dok se dispozicije razumiju kao sklonosti ili težnje koje proizlaze iz navika i oblikuju se njima.

Ključne riječi: Frank Ramsey; navika; dispozicija; pragmatizam; normativnost.

What is a Response to Wrongdoing?

Nicolas Nayfeld
Jean Moulin Lyon 3 University, France

ABSTRACT

This article starts from the assumption that to properly assess the merits of “response retributivism”, we must first clarify the nature of a response to wrongdoing and how it differs from a mere reaction. I propose a communicative account of responses, arguing that a response to wrongdoing is a distinctive form of action motivated by the wrong, whose special feature is that it is addressed to the wrongdoer, has a confrontational-communicative dimension, and needs to be identified by the wrongdoer as a response to his wrong. I argue that this definition allows us to rethink the concept of responsibility, to make progress on the debate whether there is a duty to respond or react to wrongdoing, and to refocus the discussion toward what Strawson calls "reactive practices" as opposed to "reactive attitudes".

Keywords: response retributivism; wrongdoing; communicative account; responsibility; reactive practices.

Što je odgovor na nedjelo?

Nicolas Nayfeld
Jean Moulin Lyon 3 University, Francuska

SAŽETAK

Ovaj članak polazi od pretpostavke da za ispravno vrednovanje „odgovornog retributivizma“ najprije treba razjasniti prirodu odgovora na nedjelo i razliku između odgovora i puke reakcije. Predlažem komunikativno shvaćanje odgovora, tvrdeći da je odgovor na nedjelo poseban oblik djelovanja motiviran samim nedjelom, čija je osobita značajka to što je upućen počinitelju, ima konfrontacijsko-komunikativnu dimenziju te ga počinitelj mora prepoznati kao odgovor na svoje nedjelo. Tvrdim da takvo određenje omogućuje novo promišljanje pojma odgovornosti, napredak u raspravi o tome postoji li dužnost odgovarati ili reagirati na nedjelo te preusmjeravanje rasprave na ono što Strawson naziva „reaktivnim praksama“, za razliku od „reaktivnih stavova“.

Ključne riječi: retributivizam temeljen na odgovoru; nedjelo; komunikativno shvaćanje; odgovornost; reaktivne prakse

Translated by Marko Jurjako (Rijeka)

Papers in this issue of EuJAP have been formatted by Marko Jurjako

AUTHOR GUIDELINES

Publication ethics

EuJAP subscribes to the publication principles and ethical guidelines of the Committee on Publication Ethics (COPE).

Submitted manuscripts ought to:

- be unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere;
- be approved by all co-Authors;
- contain citations and references to avoid plagiarism, self-plagiarism, and illegitimate duplication of texts, figures, etc. Moreover, Authors should obtain permission to use any third-party images, figures and the like from the respective copyright holders. The pre-reviewing process includes screening for plagiarism and self-plagiarism by means of internet browsing and software Turnitin;
- be sent exclusively electronically to the Editors (eujap@ffri.uniri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format;
- be prepared for blind refereeing: authors' names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed;
- be accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 300 words), keywords, academic affiliation and full address for correspondence including e-mail address, and, if needed, a disclosure of the Authors' potential conflict of interest that might affect the conclusions, interpretation, and evaluation of the relevant work under consideration;
- be in American or British English;
- be no longer than 9000 words, including references (for Original and Review Articles).
- be between 2000 and 5000 words, including footnotes and references (for Discussions and Critical notices)

We ask authors to submit only one manuscript at a time. A second submission by the same author is allowed only after a final decision has been made on their previously submitted manuscript.

Norms for publishing with AI

The Journal does not exclude the use of AI generated text. However, all authors (including reviewers and editors) take full responsibility for its factual accuracy and the proper acknowledgement of sources. In the acknowledgement section of your manuscript or the title page (depending on the submission/publication stage) or in other kind of reports you must identify the AI that was used, and the extent of the contribution. For instance, ChatGPT (version or the date when the AI was used).

The contribution level of the AI can be defined as follows:

- negligible – means the AI only made minor changes to the manuscript's style or grammar (this includes using AI for copyediting and similar services);
- modest – means the AI made important suggestions but was not the primary driver of the research or had an essential role in writing the manuscript;
- substantial – means the AI made several crucial suggestions that shaped the research and the manuscript could not have been completed without it.

If the contribution of the AI is "negligible", there is no requirement to mention its usage during the submission or review and publication processes. However, for any other level of contribution, it is expected that authors will report the extent of AI usage. In cases where the AI contribution is "substantial", authors, reviewers, and editors should provide a comprehensive description of the AI usage and its contributions in a narrative format.

Initial submission

When first submitting a manuscript, it is not required that the manuscript conforms to EuJAP's style guidelines. Only after a manuscript has been accepted for publication, we expect the authors to format the manuscript in accordance with EuJAP's style guidelines.

Submitting revised manuscripts

When submitting a revised manuscript, please include also a separate document where it is explained how revisions were made in response to reviewers' comments.

Policy for submitted manuscripts

If the submitted manuscript is authored by more than one person, there should be a brief explanation in the title page of the contribution of each Author with respect to the conception and design of the argument, study, etc. and writing of the paper.

To preserve the anonymous status of the review process, we prefer (but do not require) that submitted versions of manuscripts are not deposited in open access article repositories.

Policy for accepted and published manuscripts

Accepted and published versions of the manuscript can be deposited in institutional or personal repositories without an embargo period. In case of published manuscripts, a link (with DOI) to the journal's web pages and/or HRCAK should be added.

Malpractice statement

If the manuscript does not match the scope and aims of EuJAP, the Editors reserve the right to reject the manuscript without sending it out to external reviewers. Moreover, the Editors reserve the right to reject submissions that do not satisfy any of the previous conditions.

If, due to the authors' failure to inform the Editors, already published material will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue and remove the publication from EuJAP web site and the repository HRČAK.

In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors.

For additional comments, please visit our web site and read our Publication ethics statement (https://eujap.uniri.hr/publication-ethics/). To get a sense of the review process and how the referee report ought to look like, the prospective Authors are directed to visit the *For Reviewers* page on our web site (https://eujap.uniri.hr/instructions-for-reviewers/).

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations.* Cambridge: Harvard University Press.

Book with multiple authors

T: (Hirstein, Sifferd, and Fagan 2018, 100)

R: Hirstein, William, Katrina Sifferd, and Tyler Fagan. 2018. *Responsible Brains: Neuroscience, Law, and Human Culpability*. Cambridge, Massachusetts: The MIT Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, Richard. 2006. 'The Epistemic Role of Testimony: Internalist and Externalist Perspectives'. In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 77–91. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199276011.003.0004.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, Jennifer, and Ernest Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. "Normative requirements." *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)

R: Skorupski, John. 2010. “Sentimentalism: Its Scope and Limits.” *Ethical Theory and Moral Practice* 13 (2): 125–36. https://doi.org/10.1007/s10677-009-9210-6.

Article with multiple authors in a journal
T: (Churchland and Sejnowski 1990)
R: Churchland, Patricia S., and Terrence J. Sejnowski. 1990. “Neural Representation and Neural Computation.” *Philosophical Perspectives 4*. https://doi.org/10.2307/2214198

T: (Dardashti, Thébault, and Eric Winsberg 2017)
R: “Dardashti, Radin, Karim P. Y. Thébault, and Eric Winsberg. 2017. Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity.” *The British Journal for the Philosophy of Science* 68 (1): 55–89. https://doi.org/10.1093/bjps/axv010

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013. http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing ‘forthcoming’ instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, François. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, Kurt. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Copyrights

Archiving rights

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:
International:
individuals € 50
institutions € 100
Croatia:
individuals € 30
institutions € 60

Bank: Zagrebačka banka d.d. Zagreb
SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455
Only for subscribers from Croatia,
please add: “poziv na broj”: 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the *European Journal of Analytic Philosophy* are indexed and abstracted in SCOPUS, SCImago, Web of Science (Emerging Sources), The Philosopher’s Index, European Reference Index for the Humanities (ERIH PLUS), Dimensions, Directory of Open Access Journals (DOAJ), PhilPapers, Portal of Scientific Journals of Croatia (HRČAK), Sherpa Romeo (now Jisc’s open policy finder), ANVUR (Italy)