PASSION

The Journal of the European Philosophical Society for the Study of Emotion

Volume: **01**
year: **2023**
Publisher: **Open Press TiU**
ISSN: **2773-1715**

01

Table of content

Received: 15 August 2022
Accepted: 11 January 2023
Published: 15 June 2023

Editors' Introduction

Alfred Archer - Tilburg University
Heidi Maibom - University of the Basque Country/University of Cincinnati
Max Gatyas - University of Wisconsin-Superior
Joel Krueger - University of Exeter
Lucy Osler - Cardiff University

We, the editors, are delighted to introduce the new journal of the European Philosophical Society for the Study of Emotion, *Passion*. Since 2014, The European Philosophical Society for the Study of Emotions (EPSSE) has been a lively and friendly community of scholars working on philosophical issues related to emotions. The society has provided a supportive and welcoming space for scholars to develop their work, meet others working in their area and get inspired by new ideas. An important part of EPSSE's success has been its pluralistic approach. Any given EPSSE conference is likely to feature analytic philosophers of emotions, phenomenologists, ethicists, political theorists, hermeneuticians, experimental philosophers and more, discussing the emotions together in ways that cross narrow sub-disciplinary boundaries.

Like many philosophical ideas, the inspiration for EPSSE arose during a summer day in Athens. In August 2013, during the *XXIII World Congress of Philosophy*, Aaron Ben-Ze'ev and Anthony Hatzimoysis discussed the need to bring people working in the philosophy of emotions together. They envisioned a pluralistic forum that would highlight a variety of theoretical perspectives on affective phenomena, one that would encourage researchers from diverse backgrounds and in various stages of their careers to exchange ideas from across the European continent and beyond. They soon asked Angelika Krebs to help steer this effort.

Once the founding team was in place, things moved quickly. They decided to call this new society the "European Philosophical Society for the Study of Emotions," and began working on various issues such as establishing a steering committee, drafting a statement concerning EPSSE's nature and aims, and planning the inaugural conference. By the end of September 2013, the steering committee was in place, and a society statement — still on the EPSSE website — was drafted. This statement emphasizes the importance of interdisciplinary approaches to the study of emotions, the need for philosophers to contribute to this work, and affirms a special commitment to supporting early-career scholars in this endeavor. In July 2014, EPSSE's Society Statutes were approved at the first Annual Conference in Lisbon, which was a vibrant and well-attended event that attracted scholars from around the world.

Since then, EPSSE has continued to grow and flourish. In 2017, Achim Stephan took over as President of EPSSE, with Alessandra Fussi and Damian Cox as Vice-Presidents, and together they helped the society expand its

membership and achieve an even greater level of international recognition. In its 10 years of existence, EPSSE has convened 10 annual conferences and 5 workshops at several different European academic institutions.

The growing success of this community, together with the absence of an academic journal dedicated to publishing the kinds of papers that EPSSE members were writing, led the previous executive committee (led by Prof. Achim Stephan) to explore the possibility of starting a journal. The idea of a journal had been discussed for several years and had broad support amongst EPSSE members. We, the editors, took on this project when we took over as the executive committee in 2020.

Passion is a journal that aims to carry the spirit of EPSSE onto the printed (or digital) page. It is a journal dedicated to philosophical research on emotions. We aim for the journal to be open to *all* philosophical traditions. We hope to transcend the analytic and continental divide, while also conceiving of the journal as an outlet for interdisciplinary, cross-cultural, and feminist work. We especially encourage work that proposes fresh perspectives on the emotions and their interconnections with the way we live, think, act, interact, and so on. We are also open to philosophically engaged empirical work, reflecting our general attitude that *any* study of the emotions can be philosophical in nature and be philosophically impactful even when it is carried out by researchers in other fields. We aim to foster inspiring philosophical discussion amongst those passionate about emotions, from junior scholars at the start of their academic careers to experienced scholars able to draw on decades of careful study of the field. We also aim to be truly international, as reflected by our editorial board which consists of eight distinguished philosophers based in eight different countries.

After exploring various options with commercial presses, we decided instead to find a non-commercial, open access publisher to work with. This decision was largely motivated by the ever-growing and concerning controversy around profit-driven academic publishing and exorbitant open access publishing fees. These conditions place hard-working academic authors under increasing emotional and financial pressure, and effectively curtail the reach and impact of many writers' work. With such burdens disproportionately affecting early career and precariously-employed researchers, our policy at *Passion* is to resist the normalization of such practices by making our journal freely accessible and tied to a non-profit platform. This commitment reflects the wishes of the EPSSE community and our continued pledge to provide a space for novel, creative, and cutting-edge emotion research for researchers at any stage of their career. Indeed, recent developments have made the establishment of non-profit journals even more relevant and urgent. We trust that our members and other people interested in the emotions will support this movement away from for-profit journal publishing.

Open Press TiU, an open access publisher at Tilburg University (the Netherlands), were the perfect fit for our vision for the journal. Started in 2020 as part of Tilburg University's Open Science initiative, Open Press TiU aims to accelerate open access science by publishing high-quality open access academic publications. Going with Open Press TiU has the considerable advantage of being free of charge for EPSSE and of being fully open-access with no costs to the reader or the author. We are delighted to have found a publisher that shares our commitment to open access academic publishing and who have been so supportive in helping us set-up *Passion*. We are particularly grateful to Daan Rutten, Tilburg University's Open Science Coordinator, for supporting and advising us at every step of the process, helping us take *Passion* from an idea to the journal you see before you now.

We are incredibly excited to share this first issue of *Passion* with you. We believe that it showcases the diverse and vibrant kind of work being done in this thriving field of research. You will find discussions of revenge that

draw on inspirational thinkers such as Adam Smith, W. E. B. Du Bois, James Baldwin and, of course, Taylor Swift. We have rich, detailed, illuminating discussions of the role of emotions and other affective experiences in transitional justice, mental health and our interactions with robots. We are grateful to these authors for showing faith in our journal and contributing such exciting work for our opening issue.

We begin, as all academic pursuits should, with a thirst for vengeance. Our opening paper is Alice MacLachlan's 'Hell Hath No Fury: The Place of Revenge in Moral Repair'. Philosophers have tended to have a very negative view of revenge, seeing the desire for revenge as something that stands in the way of moral repair. Drawing inspiration from two films exploring revenge as a response to sexual violence, MacLachlan makes a feminist case for revenge, arguing that revenge can function as a form of moral address that takes the target of revenge seriously as a moral interlocutor and is open to their ability to change for the better. As a result, revenge can actually play an important role in bringing about moral repair.

Myisha Cherry's 'Feeling Revengeful' continues our focus on vengeance by arguing that our dominant conception of feeling revengeful lacks nuance. While people tend to assume that feeling revengeful and feelings of anger are equivalent, Cherry argues that either of these feelings can exist without the other. Moreover, acts of revenge need not involve putting the target of revenge down in some way. Instead, Cherry outlines a success model of revenge, inspired by W. E. B. Du Bois and James Baldwin, according to which being successful can be the best form of revenge against those who have tried to put you down.

Having explored the place of emotions in revenge, we turn to another important and increasingly active area of philosophical research: the role of emotions in mental health and illness. Several of our contributors have been at the forefront of this important work. Michelle Maiese's 'Are All Mental Disorders Affective Disorders?' answers this question in the affirmative. Maise argues that all mental disorders are affective in an important etiological sense, and that even overtly cognitive symptoms (e.g., disruptions of thought, language, and executive control) at least partially result from affective disturbances. Focusing on two case studies — language disturbances in schizophrenia and so-called "context blindness" in autism — Maise demonstrates how difficulties with selective attention and contextual sensitivity are common to both, and that these difficulties arise from disruptions of what she terms "affective bodily attunement" to the world.

With Matthew Ratcliffe and Louise Richardson, we consider the bounds of grief. In 'Grief over Non-Death Losses: A Phenomenological Perspective', Ratcliffe and Richardson offer a phenomenological analysis of grief beyond the context of the death of a loved one. They defend a broad conception of grief that encompasses losses other than bereavement, including loss of a relationship, loss of a job, and illness. Taking as their core case study instances of grief over involuntary childlessness, they draw attention to the way in which grief over a bereavement and other experiences of loss can share an experienced rupture of possibilities that are significant, even central, to one's identity and the structure of one's world.

From thinking about cases of personal loss and tragedy, we turn towards the role of emotions at the societal level. Emanuela Ceva and Sara Protasi's 'Framing the Role of Envy in Transitional Justice' examines the way that envy can impact upon attempts to bring about justice in the wake of large-scale human rights abuses. Through analyzing both the harmful and destructive role that envy can play and the potential fruitful and constructive effects of envy, Ceva and Protasi develop a conceptual framework for investigating the role of envy in intergroup conflict. This framework provides an important starting point for thinking about how to push these sites of conflict in more collaborative directions.

In ‘Affective Responses to Embodied Intelligence. The Test-Cases of Spot, Kaspar, and Zeno,’ Fussi examines the diversity of human affective reactions to robots using three different types of robots. Some robots elicit the famous Uncanny Valley effect, whereas others do not, but instead evoke fear due to concern about government overreach. It is especially intriguing for emotion researchers to follow the development of robots meant to improve the communication between children with autism spectrum disorder and neurotypical adults. These robots have been programmed with highly rote and simple emotional expressions in order to make it easier for such children to recognize emotion. However, as Fussi points out, given the remarkable variation in emotion expression in humans, wide use of such robots might actually have the opposite effect.

Both individually and together these authors showcase the exciting, cutting-edge, and engaged philosophical work on emotions that EPSSE aims to support and promote. We hope you enjoy this issue and feel inspired to contribute your own work in the future. By doing so you can help to make *Passion* a vibrant forum for philosophical work on emotions.

Received: 28 October 2022
Accepted: 2 November 2022
Published: 15 June 2023

Hell Hath No Fury:
The Place of Revenge in Moral Repair

Alice MacLachlan - York University (Canada)

Abstract

Revenge is a powerful word. It can conjure up the scheming, embittered individual, plotting the downfall of his enemies well beyond reason and morality—or, more seriously, tragic cycles of violence and blood vendettas, spiralling into entrenched civil conflict over generations. Philosophers have argued that the consequences and the moral psychology of revenge mean it is incompatible with—even antithetical to—any plausible conception of moral repair. In this paper I challenge that incompatibility by suggesting that, in contexts of unresponsive and imperfect institutional justice, appropriate acts of vengeance may both demand accountability and express solidarity, thus contributing to repair. Drawing on my own past work on revenge as *moral address* (MacLachlan 2016) and Peter French's conception of the *virtuous avenger* (2001)—as well as two feminist revenge films—I sketch a feminist approach to virtuous third-party vengeance as a starting point for broader, more reparative understandings of legitimate personal interventions after wrongdoing.

There are only three possible endings—aren't there?—to any story: revenge, tragedy,
or forgiveness. That's it. All stories end like that.
(Jeanette Winterson 1997)

I can tell you how it ends
Don't get sad, get even
So on the weekends
I don't dress for friends
Lately I've been dressing for revenge.
(Taylor Swift 2022)

1. Introduction[1]

In October 2017, the hashtag #MeToo swept across social media. Women around the world shared their experiences of being sexually harassed and assaulted after actor Alyssa Milano invited them to identify themselves publicly to demonstrate how pervasive sexual violence is. While the #MeToo movement, founded by activist Tarana Burke, had been operating for over a decade, 2017 became the year of the public hashtag and—for public figures—the year of reckoning. High-profile men in the Hollywood entertainment industry had charges pressed against them and several, most famously producer Harvey Weinstein, were found guilty. #MeToo ricocheted through global industries—including politics, education, technology, sports, government, finance, and the church—in at least eighty-five countries (Wikipedia 2022). The hashtag has maintained a place in public discourse since.

The power and reach of the #MeToo movement were unprecedented, causing a cultural shift in understandings of sexual violence and assault. Perhaps unsurprisingly, within three months it had provoked a forceful and lasting backlash. By early 2018, news headlines began to describe #MeToo as a "witch hunt" (Livsey 2018), "vigilante justice" (Hall 2018), "vigilantism" (Levick 2019), "mob rule" (Sharf 2018), and, crucially, nothing more than "revenge" (Branco 2018)—language that is still regularly used to discredit the movement's aims and tactics. Interestingly, much of this criticism did not grapple with the admittedly thorny issue of false accusations; the charge that #MeToo is vengeful vigilante mob justice went beyond worries about innocent casualties. It seemed intended to hold whether or not the targets were actually guilty of harassment and rape. Revenge is just wrong, it was implied, even when you get it right.

That the language of vigilante vengeance was seen as sufficient to condemn the movement highlights just how strong cultural and moral aversions to revenge can be. Revenge is a powerful word. It can conjure up the scheming, embittered individual, plotting the downfall of his enemies well beyond reason and morality—or, more seriously, tragic cycles of violence and blood vendettas, spiralling into entrenched civil conflict over generations. Vengeful people are typically seen as sadistic (Chester and DeWall 2017), malicious (Uniacke 2000), reactive (Cherry 2021), and even narcissistic (Schumann and Ross 2010). They may well be petty and spiteful. Indeed, most philosophers who take up the topic have argued that the consequences and the moral psychology of revenge mean it is incompatible with—even antithetical to—any plausible conception of moral repair after wrongdoing (Arendt [1958] 1998, Nozick 1981, Uniacke 2000, Govier 2002, Cherry 2021).

I wish to challenge that presumed incompatibility. While I don't actually believe that the practices of #MeToo (i.e., publicly recounting one's experience of sexual assault and, potentially, naming one's attacker when doing so) count as revenge, I am interested in the *power* of the accusation that they were: the idea that qualifying as revenge automatically delegitimises a response to wrongdoing. Instead, I argue, in contexts of unresponsive and imperfect institutional injustice, appropriate acts of vengeance might both create accountability and express solidarity with victims, thus contributing to crucial aspects of moral repair. In other words, my interest in revenge is provoked by, but ultimately not centred on, practices of #MeToo activism—though the context of my analysis will be similar (sexual violence and misconduct).

1 I am deeply grateful to audiences at the Atlantic Regional Philosophical Association (Halifax, Canada, 2007) and "Let's Talk about Revenge! Retributive Emotions, Justice, and Moral Repair" (Essen, Germany, 2022) for their thoughtful engagement with previous versions of this work. My deep thanks also to Alfred Archer and to an anonymous reviewer for their helpful comments and criticism, and to Myisha Cherry for helping me work through some of the thornier aspects of vengeful emotion. Finally, I would like to thank Catherine Clune-Taylor and Deborah Finding for encouraging me to think more deeply about *Promising Young Woman*, Elliot Page in *Hard Candy* for initially inspiring my thoughts about feminist vigilantism, and David Wyatt for starting all of this thirty-five years ago when he read me *The Princess Bride*.

My argument in defence of revenge takes the following form: first, I examine the philosophical case against revenge, and argue that both categories of objection can potentially be answered. I then turn to my previous work on revenge as a form of *moral address* (MacLachlan 2016) and to Peter French's argument in favour of recognising potentially *virtuous avengers* (French 2001). Third, I illustrate the potential of virtuous vengeance as a response to sexual violence (as well as its pitfalls) by considering two films featuring feminist avengers of sexual assault: *Hard Candy* (2005) and *Promising Young Woman* (2020). Drawing on these narratives, I supplement French's notion of virtuous vengeance with what I call *reparative* revenge, focusing especially on the value of interpersonal accountability and victim solidarity. I recognise the significant risks and complexities of any campaign of vengeance, but I suggest that acknowledging its value broadens our understanding of what qualifies as an appropriate personal intervention in the aftermath of wrongdoing, especially in contexts of institutional failure.

2. The Case against Revenge

> Revenge is a kind of wild justice, which the more man's nature runs to, the more ought law to weed it out.
> (Francis Bacon [1625] 1986)

Philosophical arguments against revenge tend to fall into one of two categories. I call these objections (i) escalating consequences and (ii) destructive motivations—in particular, seeking satisfaction in the suffering of another. While the latter objection focuses on affective states internal to the act of revenge (the emotions and aims of the revenger), the first concerns itself with the risks of a *culture of revenge*. Robert Nozick articulates one source of this risk in *Philosophical Explanations* (1981), when he outlines the differences between revenge and retribution. Among them, he argues, is that revenge—unlike retribution—"sets no limits" on what can count as payback and, further, has no generality requirement. Whether or not something warrants revenge depends on the revenger's attitude to the wrong and wrongdoer at the time, and the revenger is not thereby committed to the principle that similar acts would warrant similar responses; in particular, she is not committed to the idea that she might similarly be accountable. In other words, practices of revenge are unregulated and irregular, with no internal brakes in either a single case or across cases (367–68). Hannah Arendt also views revenge as contrary to agency and choice—not because it is unpredictable and chaotic, but rather because it is inevitable: "far from putting an end to the consequences of the first misdeed, everybody remains bound to the process, permitting the chain reaction contained in every action to take its unhindered course" ([1958] 1998, 240–41). In a society where everyone remains wholly committed to revenge, one act of wrongful harm could set off a chain reaction that has no end. In other words, while a singular act of revenge may do no real significant harm, the risks of a culture that promotes or even allows for revenge are too great to allow ourselves to contemplate even a one-off instance.

Yet, as Jon Elster points out, social practices of revenge "can be subject to very elaborate norms" (1990, 870). These internal standards include norms of agency (who can take revenge for a particular harm), proportionality (what counts as appropriate revenge), and target (who can legitimately be harmed in the act of vengeance), and are bound up in understandings of shame, honour, and kinship. Because these norms are both socially understood and enforced, on one hand, and complex enough to require careful judgment and consideration, on the other, they are unlikely to allow for either the wild and unpredictable sprees Nozick predicts or the kind of automaticity Arendt fears. In other words, widespread practices of revenge may well protect against the variability of one-off *acts* of revenge.

Even without a culturally established set of norms, there is no reason to think that acts of revenge are necessarily open-ended. Proportionality (*lex talionis* or "an eye for an eye") has played a crucial role in concepts of retaliation as far back as Hammurabi's Code; while individuals may disagree about the precise magnitude of a proportional response, most people see some responses as proportionate and not others. As Suzanne Uniacke—herself a critic of revenge—notes, "successful retaliation within limits can be satisfying and some people know when to stop. Revenge can be conducted in secret and can be confined to the relevant parties and remain an isolated event" (2000, 64).

Indeed, acts of revenge need not be violent or violate the rights of the target. While it is tempting to envision duels and devastation, sometimes the most satisfying acts of revenge are a well-placed cutting remark, a skipped invitation, a down-vote at a meeting, or the decision to cc the right people when replying to an embarrassing email. Robert Solomon is quite explicit on this point: we live "in a tit-for-tat world. We are all moral accountants, even if the bookkeeping varies considerably" (1999, 126)—in most cases being "in the red" in such accounting does not lead to actual bloodshed. Fabian Bernhardt pushes this point further, arguing the tendency of early modern political philosophers to characterise revenge as uncontrolled, dysfunctional, and excessively violent may have something to do with the desire to mark a sharp boundary between (public) punishment and (private) revenge, to "provide normative ground for the rule of [the] state and its monopoly on the use of retributive violence" (2020, 503).

The second set of philosophical objections to revenge focuses on the emotions and motivations that lead to it: rage, resentment, malice, hatred, and the desire for payback or to "lower" the target in some crucial way. Revenge is morally problematic because is motivated and given a rationale by our uglier, "vindictive" passions.

Uniacke argues that revenge is typically malicious and vindictive (2000, 62)—seeking out evil or harm to another person and deriving pleasure from that harm—but not always. Rather, "the emotion that gives rise to the desire for revenge is resentment: bitter feelings about an injury sustained" (63). Not only injuries can prompt such bitter feelings. The desire for revenge may be prompted by a slight, affront, or threat to our reputation or self-esteem, even as we recognise the provoking action may have been accidental or unavoidable ("she didn't mean to leave me out of her thank-you list, sure—but she should still pay for that public humiliation"). Uniacke understands resentment as both pain at the offence and the sense that the pain can only be alleviated by striking back, returning like for like, or by contemplating fantasies of doing so. The tendency of such bitterness to *become* malicious is certainly concerning, but Uniacke condemns even acts of revenge that are "non-vindictive or relatively harmless" (64). Indeed, for Uniacke, the distinctive emotional motivation and rationale for revenge is what makes it wrong:

> Some acts of revenge, although morally inappropriate *qua* acts of revenge, can be independently justified. A book may deserve an unfavorable review on its merits even though the reviewer's motive is vindictive. (2000, 68)

Uniacke acknowledges our attitude toward revenge and revenge-takers is more mixed than other emotions associated with pleasure at the suffering of others (e.g. pure malice or *schadenfreude*), due to our sympathy with their plight as victims and the intuitive appeal of payback as a kind of karmic balance. She also acknowledges the naturalness of resentment as a response to injury, but she concludes that this emotional response "cannot morally justify reprisal in kind" (68).

Myisha Cherry also connects revenge to resentment, critiquing what she calls "ressentiment rage" as morally and politically problematic because its aim is revenge. Cherry's context of analysis is rage arising from racism and the disempowerment of racialised groups; she writes, "the outraged wishes for revenge as payback for the racial group taking away his group's power. And he may wish for and even cause physical, mental, or status harm as a result" (2021, 19). For Cherry, the central problem with this vengeful emotion is not that it is angry, but that its anger is reactive; in needing to "lower" the other or settle the score, the revenge-seeker continues to centre the target's agency and not their own. Vengeful rage is passive and self-defeating: its continued focus on its target only feeds feelings of envy and insecurity, rather than bolstering the agency of the oppressed.

In later work, Cherry expands her analysis of vengeful emotions, arguing that anger is neither necessary nor sufficient for what she calls "feeling vengeful." Instead, the passions of revenge (which Cherry calls "vindictive passions") include shame, contempt, resentment, disgust, and love of justice, emotions that have been overlooked in the philosophical focus on anger and hatred (2022, 5–6). Cherry's moral psychology of revenge includes the revenger's *plans* to harm the target, the vindictive passions that fuel these plans, the pleasures that come with either completing the plan or contemplating its completion (dreaming of revenge), and the beliefs that this kind of harm is instrumental for righting wrongs or gaining recognition (ibid.).

The emotional case against revenge is compelling, admittedly, but it also risks a certain circularity unless we can independently establish the wrongness of *taking* revenge. After all, the vindictive passions are considered unpleasant precisely because of their association with the desire for revenge. The same goes for the rest of the moral psychology of revenge: planning to take revenge, dreaming of those plans, and deriving pleasure from that dreaming—these are wrong only insofar as the plans themselves are wrong, i.e., insofar as revenge is wrong. Trudy Govier claims that the desire for revenge is an evil desire, and that to act on it is "to indulge and cultivate something evil in ourselves," resulting in our moral diminishment (2002, 13). What makes the desire for revenge, and the moral psychology that develops around this desire, so evil? As Govier asks:

> Suppose it were non-obsessive, non-violent, kept proportional and within bounds, applied to those who really were the wrongdoers and not inflicted on innocent third parties, and satisfying in the end. In such a case, could revenge be right? In other words, is there anything wrong with the desire for revenge or the quest for revenge as such, considered apart from its consequences? (11)

For Govier, Uniacke, and others, the answer is that the revenger desires—as a kind of core aim—that another person suffer for my satisfaction.[2] In feeling vengeful, I need them to feel bad so that I can feel better. Their suffering may not include actual physical violence and may not violate their rights, but it still treats them as a means to my end. This offends against respect for persons and human dignity, of course—and seems especially pernicious when what I am using is their (non-consensual, and presumably unwanted) *harm* for the end of my *satisfaction*. Insofar as the vengeful passions are identifiable by this core desire and the psychology of revenge built around it as either project or fantasy, these passions and this psychology are immoral, and taint whatever actions are taken as a result.

2 As Govier says, "when we seek revenge, we do so in order to take pleasure in the fact that *the offender has been made to suffer* and *it is we who have brought this about*" (2002, 13, italics added). In Uniacke's words, "the satisfaction that someone who indulges in revenge seeks or derives from another person's suffering is a matter of retaliation for an injury" (2000, 67).

3. Moral Address and the Virtuous Avenger

At first glance, this is a damning criticism of revenge. Even if things don't escalate, even if everything remains non-violent, even if no one's rights are violated, even if malice and spite are kept in check – at its core, these philosophers argue, the intent to enact revenge is always disrespectful of human dignity. Revenge treats its target as an object to be used, *through the infliction of harm* (however minor), and not as a person to be engaged. Revenge fails to acknowledge the extent to which we are *all* guilty of inadvertent and intentional slights and harms to others (not to mention more serious injuries . . .)—and we do not, and should not, lose access to our status as persons with dignity as a result. The moral bookkeeping of revenge leaves little room for grace. But the worry goes much further. Learning to see those who wrong us as mere objects on which we are *allowed* to inflict vindictive satisfaction for our own emotional needs cannot help but contribute to, among other social ills, deeply unjust and inhumane attitudes towards incarcerated persons. Anything that comes to see others as crucially *less*, even disposable, will work against attempts at moral repair following wrongdoing.

But is this what revenge does? A defence of the practice requires that we look more closely at what it means to take revenge. In his 2006 paper "Revenge," Robert Stainton meticulously itemises the various conditions on revenge and, specifically, how an act that might seem to be revenge can fail as such, depending on the internal states of the revenge-taker and target. After reviewing several cases where one person, A, aims to take revenge on another, B, and fails, Stainton concludes that "the ideal cases are those in which B sees the *connection* between harm H_A which she did, and harm HB which is now done to her. And that is likely to require B to realise that it was A who brought about H_B" (2006, 17). This is quite plausible. If B never realises that they have not just suffered some unfortunate accident of fate but have, in fact, been brought *low* by A, because of B's past actions, then there is some sense in which A's vengeful feelings will not be satisfied by the act. Put crudely, "an eye for an eye" is not satisfied if the original eye-taker suffers an unfortunate accident with a fork in the face shortly after their misdeed.

But already, this explanation pulls against the idea that revenge treats the target as a mere means to my satisfaction. We do not require this level of recognition from objects; such engagement is appropriate to fellow subjects. The desire for revenge isn't just a desire to shift our present feelings to the target (e.g., to make our pain and suffering now theirs) but also, to share our plans (that they are being hurt *by* us, intentionally and because of their prior actions) and, ideally, our beliefs (that they are not beyond our reach, that they have brought this on themselves, perhaps even that they *deserve* what has come to them). Adam Smith, writing in *The Theory of Moral Sentiments*, captures the persuasive element of vengeful resentment beautifully:

> Resentment would prompt us to desire, not only that he should be punished, but that he should be punished by our means, and upon account of that particular injury which he had done to us. Resentment cannot be fully gratified, unless the offender is not only made to grieve in his turn, but to grieve for that particular wrong which we have suffered from him. He must be made to repent and be sorry for this very action. ([1759] 1976, II.i.I.6)

Here we see the vindictive picture captured in contemporary philosophical accounts: the desire that the wrongdoer experience pain ("be punished"), and specifically that we be the agent of that pain ("by our means") and that the wrongdoer's pain be *tied* to ours in a way that both we and he are aware of ("upon account of that particularly injury"). Moreover, Smith rightly identifies the ways in which we want the offender to transform: "he must be made to repent and be sorry." As I have argued in past work on the topic (2016, 140), the aim

of revenge is not only harm to another, but *transformational* harm. Smith is describing a fantasy of forcible moral persuasion; the target of resentment is not merely an instrument for its satisfaction, but audience for its message. The revenger wants the target not only to feel differently, but to think differently. Revenge is, among other things, an act of communication (admittedly a fairly forcible, unilateral communication in most cases)—what I have called a form of *moral address*.[3]

The idea of moral address opens up the moral potentialities of revenge, recasting it as part of a moral communication between revenger and target, or—in their original roles—victim and wrongdoer. Recognising the communicative dimension of revenge also widens our understanding of its emotional dimension; someone might be desperate to initiate this kind of moral conversation not because they are overwhelmed by malice, but because they are overwhelmed by love or grief. I might fantasise about aggressively confronting the negligent driver who killed a family member not because I am filled with malice, but because I can't bear the idea that my loved one didn't matter in their eyes, might have meant less to them than the text they were answering. Similarly, suppose a friend's assault was never successfully prosecuted, whether because of local failures or—equally likely—some structural or systemic bias. I might entertain fantasies of revenge out of an intensive form of third-person indignation—the need for the wrongdoer to recognise that at least one person knows what they did, knows that it was wrong, and intends to hold them accountable.[4]

Yet while the idea of moral address opens up conceptual space for permissible—even morally productive—acts of revenge, it provides us little guidance for distinguishing acts of revenge likely to contribute to the righting of wrongs from those which may exacerbate the effects of wrongdoing by reigniting hostilities, opening old wounds, and so on. There are many permissible actions that are nevertheless profoundly unhelpful and thus problematic in the delicate aftermath of wrongdoing. Margaret Walker invites us to consider the moral work of this aftermath as moral repair: "the task of restoring or stabilising—and in some cases creating—the basic elements that sustain human beings in a recognisably moral relationship" (2006, 23). Most of the time, moral repair will entail holding wrongdoers accountable and responsible, addressing the harms themselves and acknowledging victims and survivors, engaging in rituals and practices that establish or re-establish trust (e.g. apologies and rituals of forgiveness), asserting or reasserting appropriate normative horizons, and—where possible and appropriate—building or rebuilding appropriately moral relationships among those implicated in the wrong.[5] How might appropriately communicative and constrained acts of revenge fit into this reparative picture? To answer this question, I turn to a second philosophical defence of vengeance, offered by Peter French.

3 One of the best examples of revenge as moral address can be found in the 1973 William Goldman novel and 1987 film *The Princess Bride*: the Spanish fencer Inigo Montoya has spent his life seeking his father's murderer to enact vengeance and practicing the speech he will give when he confronts them. "Hello. My Name is Inigo Montoya. You killed my father. Prepare to die." I have argued that, if Montoya had failed to identify and address his target, identify himself, name the wrong, and connect it to the suffering he was about to inflict, he would not have succeeded in taking revenge (despite duelling with and ultimately killing his target). See MacLachlan 2016 for more details.

4 Indeed, the longed-for conversation need not always be directly with one's wrongdoer. Sometimes resentment and the desire for revenge linger as a form of protest against others (even a whole society) who have simply *moved on*, without wholly attending and responding to the depths of the wrong. Jean Améry describes the loneliness of being out of time in relation to others by refusing to look away from the horrors of his own torture or the genocidal reality of the Nazi Holocaust. Améry contemplates the fantasy of revenge as a form of *fellowship* with his persecutor: "when the SS-man Wajs stood before the firing squad, he experienced the moral truth of his crimes. At that moment, he was with *me*—and I was no longer alone with the shovel handle. I would like to believe that at the instant of his execution he wanted exactly as much as I do to turn back time, to undo what had been done" (2009, 70).

5 Walker is clear that she does not intend moral repair to be a synonym for reconciliation. Morally right relationships can be as minimal as mutually acknowledging the other's humanity before separating entirely. But obviously the physical and psychological safety of victims must be prioritised in the establishment of *any* relationship with wrongdoers.

The careful reader will have spotted the shift in the examples of the negligent driver and the assault: in each case, the revenger is not the original primary victim of the wrongdoing. In the case of the negligent driver, the revenger is arguably a secondary victim of the injury: the driver's negligence has harmed them by taking a loved one from them. In the second case, however, the revenger is clearly an invested third party, motivated to take vengeance because of their care for the victim, their friend. And yet, it does seem that while revenge is *personal*, the injuries which I can reasonably take personally are not limited to my own. Nozick, for example, specifically notes that the revenge taker should have "special or personal tie to the victim" as a condition of revenge: "'this is because of what you did to my ___' (self, father, group, and so on)." (1981, 367).[6] Similarly Elster acknowledges that "who [is] allowed or required to exact revenge" for a certain offense is one of the complex governing norms of revenge cultures, meaning that the question is not a foregone conclusion and must be negotiated among kin or group (1990, 867).

While most still insist that the tie must be close (e.g. kinship) for the act of retaliation to be sufficiently personal to count as revenge, Peter French stands out as the exception. In *The Virtues of Vengeance* (2001), he argues that the moral value of revenge as a response to evil has been overlooked and establishes a set of conditions under which, he argues, vengeance is virtuous. Intriguingly, while French acknowledges that "revenge is typically more personal than retribution [and t]he avenger is often linked in some crucial way to the person or persons who were injured, harmed, slighted, by the target" (67), he argues that this personal connection is not necessary. Rather, what he calls the "authority" to enact vengeance derives from independent moral qualities of the revenger—in French's terminology, the avenger—and not their personal connection to the victim.[7]

French's defence of revenge sees it as an individual moral effort to correct for both the nonkarmic nature of the universe and the failures of impersonal institutions and mechanisms of justice. When both the universe and the authorities fail to respond adequately to harms and wrongdoing, revenge picks up on a basic moral intuition: that wrongdoing somehow *demands* a hostile response. For the most part French's focus is the film genre of Westerns, which he reads as both expressions of the cultural and moral values of revenge and as an illustrative context: practices of revenge arise in conditions of moral and state lawlessness. Vengeance fills the gap left by institutional failure or absence and ultimately, it becomes a way to rectify structural social inequalities and those who are vulnerable to them.

Like Elster, French argues that for revenge to be coherent as a practice, it must be norm governed and intelligible within a set of recognisable social practices; like I do, he believes that revenge is fundamentally communicative. For vengeance to be successful, the target must understand that she is suffering as a penalty for the actions that triggered the revenging behaviour. Ultimately, however, French's defence rests on the possibility of there being acts of *virtuous* vengeance; that is, conditions under which taking vengeance is the right thing to do—that is, where it displays the appropriate motives, dispositions, and emotions.

In addition to revenge being both intelligible as a cultural practice and appropriately communicated to the target (both conditions which distinguish moral address from mere satisfaction) a virtuous act of revenge must meet three conditions. First, the target must *deserve* it. This rules out both cases of mistaken identity

6 For a parallel discussion of who does or doesn't hold the prerogative to *forgive* wrongdoings to others, see Pettigrove (2009) and MacLachlan (2017).

7 French is quite clear on this point: "I believe that Nozick's second point of difference between revenge and retribution—that revenge can only be inflicted by a person with a personal tie to the victim—is wrong" (68).

or attribution, as well as non-culpable and minor slights, offences, and injuries. While French is not explicit on this point, his discussion implies a certain threshold of wrongdoing above which revenge is appropriate.[8] Closely connected to the desert condition is that of *proportionality* and fit: the penalty should, in some sense, fit the crime, both in magnitude and in quality or nature. A penalty that was significantly *less* than what was warranted might fail to register as revenge, while one that went too far would constitute a new act of aggression. Understandings of fit/proportionality will vary from context to context, and they may be open to dispute; French says avengers must be prepared to explain their choice of penalty, remarking that "the virtues of vengeance are regulated through the moral dialogue of the community" (229). Finally, the virtuous avenger must possess the *authority* to take revenge. This authority is not bestowed by the victim or the avenger's relationship to her; instead, the avenger holds authority if she acts on the right motives (for example, appropriate motives and desires), possesses the appropriate virtues of character (French suggests both inner strength and moral originality), and uses an epistemologically reliable procedure to determine her target and penalty.[9]

French's picture of the virtuous avenger fills in the opening created by the idea of revenge as moral address; his virtuous avenger is someone who, given her appropriate motives, upright character, and reliable procedure, is able to grasp or create the appropriately proportionate, fitting response to a sufficiently significant offense, ensuring that her response is directed at the deserving target, and that it communicates the moral reasons for which the penalty is inflicted. If French is right, then this avenger's actions are both textbook revenge and—unless she commits some other wrong in performing them—entirely morally acceptable. Moreover, French's emphasis on the ways in which practices of revenge fill the gaps left by inadequate institutional justice hints at how revenge might play a role in moral repair. If moral repair requires holding wrongdoers accountable and finding ways to set people back in right relationship, the kind of virtuous vengeance French has in mind could potentially play a role.

4. Two Cases: *Hard Candy* and *Promising Young Woman*

French's argument is illustrated largely by the cultural archive of classic Westerns. I want to turn to two other films, belonging to what is sometimes (unfortunately) named the rape-revenge genre. I choose these films for several reasons. First, they return us to the domain which first motivated my interest in this topic: namely, the legitimacy of personal retributive responses to sexual violence in the absence of institutional justice. Second, and importantly, the differences between the two revenge-taking protagonists illustrate crucial elements of virtuous vengeance that French overlooks—and which I argue redirect our attention away from virtuous vengeance, and towards *reparative* revenge. These films are David Slade's *Hard Candy* (2005) and Emerald Fennell's *Promising Young Woman* (2020).

Hard Candy was billed as a cat-and-mouse game between a fourteen-year-old vigilante, Hayley, and Jeff, a man she suspects of being a sexual predator. The film opens with their online flirtation, a tentative meet-up at a

8 Presumably there is both cultural and individual variation here. While few people would see failing to replace the milk in the fridge as a revenge-worthy offence, some would see forgetting a birthday or important anniversary might as revenge-worthy, while for others revenge only becomes a question in cases of significant betrayal and deceit (e.g., keeping a secret second family or cleaning out a shared savings account).

9 This might seem redundant, given the desert and proportionality condition, but for the avenger to act virtuously, it can't be a coincidence that her actions are directed at the right target, in the right way.

coffee shop, and his studiously casual invitation to come back to his place.[10] Yet the audience soon learns that Hayley is wholly in control of this interaction; indeed, she already knows that Jeff colluded with another man (Aaron) to rape, photograph, and kill a local girl, because this is her second vengeful confrontation. She has been tracking and baiting both of them and, moreover, one of the first things she does is to locate the pictures of Donna that Jeff has locked in his safe. In other words, we are shown Hayley has an epistemologically reliable method for determining her target.

Over the course of the film, things between Hayley and Jeff escalate radically. When he tries to give her alcohol, she insists on making the drinks herself and drugs him (at which point she locates the evidence and ties him to a chair). When he manages to retaliate and overpower her, threatening her with the gun, Hayley manages to choke him with plastic wrap from behind. In the next scene, Hayley falsely persuades Jeff that she is castrating him (by putting ice on his genitals) which is traumatic for Jeff. She walks away, he attacks her with a scalpel, and she retaliates with a stun gun. In the final act, Hayley persuades Jeff to kill himself, by telling him that if he doesn't, she will expose his secrets; killing himself is the only way to keep them. Jeff attempts to bargain and plead, offering to tell her the other man's name, at which point Hayley reveals that she already knows it, and Aaron is already dead. Jeff jumps off the roof with a noose around his neck and Hayley leaves the evidence for the police to find, despite her promise, and walks away.

In some ways, Hayley most resembles the classic avenger that French envisions. Her connection to the original victim is minimal—they are both young girls vulnerable to sexual predation, but it's not clear that they even knew each other. Hayley acts on her own sense of indignation and rage at the impunity of predators and the apathy of everyone else. Her chosen revenge scheme is both communicative and, arguably, fitting. Her language and actions mimic those of Jeff and other predators, and she narrates this to him (and the audience): tracking her target online, interfering with his drink, overpowering him when he passes out, and assaulting his genitals against his explicit wishes while filming his assault. Jeff's end resembles that of the original victim, Donna—except, unlike Donna, Jeff seems to be given some agency. Hayley's revenge scheme depends entirely on Jeff making culpable choices at each stage: bringing a fourteen-year-old to his house, offering her alcohol, first threatening her, and then attacking her with a gun and scalpel. Each escalation in violence is predicated on his initial decision to attack.

Yet the agency given to Hayley's target, Jeff, is largely illusory. Although he consistently acts badly, bringing on the worse option, it's not clear that if Jeff had experienced genuine contrition at any point (including the willingness to turn himself into authorities) Hayley's plans would have changed. Her communication is one-sided, and while it's true the film's audience would likely rather not hear much more of Jeff's perspective, this still makes for a profoundly unilateral address: moral command rather than moral persuasion. As I have previously argued, morally problematic "revenge fails to participate in a moral dialogue . . . it always aims to shut that dialogue down, ending the moral conversation" (2016, 144) because the revenger cannot be open to her target's perspective. Hayley's ability to do what she does depends on her continuing to believe in the righteousness of her cause, and that requires silencing her interlocutor, rather than being open to any reciprocal persuasion. Again, the audience is not expected to mourn Jeff's silence, but this attitude carries the risk of disrespect and, worse, the idea that some people need never be heard from and can simply be silenced and then disposed of.

10 While the protagonist, Hayley, is played on screen by Elliot Page, my discussion concerns the character and not the actor. Thus, I will use "she" and "her" as Hayley's—and not Page's—pronouns.

Even more worrying, perhaps, is the vanishingly small role played by Jeff's victim, Donna, in the film and in Hayley's machinations. While Hayley may address Jeff, in some ways she *uses* Donna to do so, even leaving out explicit photos of Donna's rape and death in a sordid display at the end of the movie. Donna becomes the means by which Hayley achieves her goal of punishing Jeff for his predation: while the viewer is persuaded of Hayley's genuine and deeply sincere anger and indignation at Jeff and those like him, we see almost no concern—or even much thought—for Donna and those like her. Hayley may achieve accountability for Jeff, but her actions fail to prioritise the perspective and value of his victim(s).

Hayley's disconnect from the victim she claims to avenge contrasts sharply with Cassie, the protagonist of *Promising Young Woman* (2020). Cassie is entirely consumed by the loss of her best friend, Nina, who committed suicide after a med school classmate of theirs, Al, raped her. Cassie is clearly grieving; she has dropped out of med school, works at a coffee shop, and lives with her parents; at first, she seems to have no plans and no direction. She spends her nights pretending to be drunk in clubs and bars so that men will take her home and try to take advantage of her, at which point she reveals she's completely sober, "thrusting them into a jarring moment of self-reflection," in the words of critic Aisha Harris (2021). But, in addition to her habitual practices of moral confrontation, Cassie has also embarked on a comprehensive program of revenge on Al and everyone else she holds responsible for Nina's suicide. This includes Madison, a classmate who blamed Nina for her own rape because she had gotten drunk; Elizabeth Walker, the college Dean who dismissed Nina's case for lack of evidence and then celebrated Al as a visiting alumnus; and Jordan Green, Al's lawyer who harassed Nina into dropping the charges against him. Nina's rapist, Al, is Cassie's final target.

Like Hayley, Cassie has planned her revenge extraordinarily carefully. Her targets are well-researched, and her plans for each are both proportionate to their degree of responsibility and fitting modes of communicating the moral message Cassie needs them to hear. For Dean Walker, who insisted there was no reason to doubt boys of character, Cassie highlights the inconsistency between Walker's careless treatment of the young women students in her care, on one hand, and Walker's fierce protectiveness of her own young daughter, on the other. She deceives the Dean into thinking that Amber, her daughter, has been alone in a dorm room with drunk male students for several hours, and now can't be reached; Walker finds herself panicked, desperate, and terrified—utterly at odds with her earlier casual reassurance that, after all, these are "good boys of character." The deceit is the only real harm Cassie inflicts on Walker; in fact, Amber has been persuaded to wait safely in a diner, and Cassie has hidden her phone. For Madison, who engaged in victim-blaming and thought Nina should have taken more responsibility for her drinking, Cassie roofies her drink at lunch, and then pays a gay man to take Madison to a hotel room so that Madison can experience waking up wondering if she's been raped. For Al, who treated Nina like dirt and put her behind him as he moved on with his life, Cassie escalates to violence: her plan is to carve Nina's name into his chest while he is handcuffed to a bed at his bachelor party. Cassie will leave Nina's mark on his body and psyche, as he violently left his mark on Nina's.

The audience never learns exactly what Cassie's plan is for Jordan Green, the lawyer who harassed her friend. When she arrives at his house, she learns he is on leave after having a nervous breakdown following Nina's suicide. He is deeply guilty and remorseful. Instead of harming him, Cassie forgives Jordan and ends up comforting him. She doesn't need to take further action because her revenge has been satisfied, revealing that her aim had always been accountability through suffering—in Jordan's case, the suffering of his own remorse—and not her targets' suffering for its own sake.

Cassie's actions, like Hayley's, depend on the reactions and responses of her interlocutors, but, unlike Hayley, her plans for revenge are open to revision in light of remorse and contrition. Cassie's plans are also almost entirely non-violent (with the exception of Al). Her revenge aims to *shock* the target more than harm them, to shake them into moral transformation. This is true both of her purposeful plans for revenge on behalf of Nina, and her more aimless—though arguably vengeful—practice of disturbing would-be rapists by suddenly revealing her sobriety. In both cases, Cassie's vengeful strategies are in line with philosophers like Macalester Bell (2013), Ami Harbin (2016), and Imke von Maur (2022) who argue that experiences of cognitive and affective disorientation, disturbance, and disruption can be valuable for moral agency, shaking our complacency and settled, habitual expectations, inviting us to recognise the reasons we have for doing and thinking differently, forcing us to ask: "how can I go on?" In other words, disorienting shocks of the kind that Cassie inflicts can play a role in holding wrongdoers accountable and in having them internalise that accountability.

Cassie enacts revenge as a way of forcing her targets to confront their wrongdoing (Al), their hypocrisy (Madison and Dean Walker), or their cowardice (Ryan, the former classmate Cassie was in love with until she learnt he was a silent bystander to Nina's rape, is forced to silently witness his friend Al's downfall in a parallel bystander role; in doing so, she makes Ryan confront his own weakness).[11] In some ways, her campaign of vengeance is less a moral conversation she has with her targets than a conversation she forces them to have with themselves. At the same time, her response to Green's contrite devastation reveals her openness to genuine moral conversation. Additionally, the audience learns that for Cassie, revenge is a last resort, a way to have the moral conversation that Al's moral community (his friends and classmates) and at least two institutions (the university and the courts) refused to have with him.

Finally, she never loses sight of Nina as the heart of her vengeance. Cassie is deeply vengeful and at times she is intensely angry, but the emotional source of her vengeance is always inflected by her abiding love and grief for Nina, represented by the half-heart-shaped necklace with Nina's name Cassie wears when she sets off to confront Al.

5. Revenge and Moral Repair

I take Cassie and Hayley to be both illustrative of and a challenge to French's ideal of the virtuous avenger. While French draws his moral picture from Westerns, a genre premised on the absence of law and order ("the Wild West"), I have turned to a very different genre—albeit one that is also premised on the absence of functioning institutional injustice. Both Hayley and Cassie live in a "wild west" of sexual violence, where predators and rapists are able to act with impunity and without fear of accountability. They both enact programmes of revenge that are, on one hand, targeted at specific wrongdoers and which, on the other hand, function as censure of the lawless state they find themselves. At one point Hayley tells Jeff directly: "I'm every little girl you ever watched, touched, screwed, killed." Cassie takes this symbolic status further; her nightly habits turn her into an everywoman of sorts: the universal drunk girl at the bar, as a silent symbol of vulnerability. For both women, revenge is also an act of protest.

Cassie's and Hayley's aims draw our attention to a role for revenge largely overlooked in the philosophical literature. Insofar as the avenger steps up when no one else will, their actions—and their sheer willingness

11 Since *Promising Young Woman* is a relatively recent film and relies on some significant plot twists, I try to remain as vague as possible about the ending.

to act—function as an implicit rebuke to the institutions and authorities that failed to intervene, and they highlight how the risks of acting outside the law are outweighed by the frustrations and pointlessness of working within it. When the wrongdoings in question are facilitated and enabled by broader issues of injustice and oppression, such protest contributes to processes of moral repair insofar as it draws our attention to conditions of mistrust and what needs to change. Dean Walker has been forced to confront her willingness to overlook sexual violence on campus; Madison will hesitate before condemning other women.

The contrast between Hayley and Cassie also points to the limits of French's account. Hayley most closely resembles the virtuous avenger as French describes them. Yet, above, I identified what I took to be two moral weaknesses in Hayley's approach to vengeance unaccounted for in French's framework: first is the unilateral nature of her revenge-taking. While Hayley pretends to engage Jeff in moral conversation—letting him tell her about his own childhood abuse, pretending to consider a bargain of death for secrecy—ultimately her communication is more command than conversation. This is perhaps best illustrated in the moments before Jeff jumps off the roof with the noose around his neck. Second, Hayley fails to stand in appropriate moral relationship to the victims whose wrongs she aimed to avenge. Donna—the girl whom Jeff killed—only features peripherally in Hayley's decision-making, but is left exposed (literally, in photographs) in her wake. Hayley may be avenging Donna's death, but her vengeance is something she does *to* Jeff, and not *for* Donna. She may be relevantly similar to Donna, and take revenge on her behalf, but there is a very real sense in which Hayley fails to stand in solidarity with Donna, precisely because she lacks an appropriate relationship of care or concern.

Cassie's actions, by contrast, demonstrate both receptivity to her targets as interlocutors and openness to their ability to change for the better, as shown in her interaction with Jordan. Similarly, Cassie consistently engages in reparative rather than retributive vengeance: the penalties she inflicts are specifically selected for their ability to disrupt her target's perception of their own actions and agency, putting them in a position to *see* themselves as responsible and accountable, thus inviting them into a moral conversation via their own remorse and atonement. If one of the moral risks of revenge is its tendency to function like a moral mic drop, ending the conversation exactly where the revenge taker wishes, Cassie exacts her revenge in a way that minimises that moral risk by remaining vulnerable to further interjections from others. Indeed, Madison shows up later in the movie, to confirm that Cassie's actions had their intended effect by helping—and thus endorsing—Cassie's cause. Madison then angrily tells Cassie never to contact her again, revealing to Cassie—and the audience—what it *felt* like to be on the receiving end of Cassie's moral address, even if she was persuaded by it. The audience is also allowed to see how much Madison's distress—and Green's contrition—leave Cassie shaken.

Second, Cassie's consistent focus on Nina and their relationship challenges both French's insistence that the avenger's authority has nothing to do with the victim, and philosophical accounts that locate the desire for revenge in resentment, malice, and hatred. While Cassie is undoubtedly angry, her governing emotions are grief and love, both of which play a role in *how* she avenges Nina; her recognisable rage and anger never overpower these. Nina is so present in Cassie's emotions and decision making that her final cumulative act is to carve Nina's name into the flesh of her rapist, insisting that he remember Nina as vividly as Cassie does. Furthermore, Cassie's only moments of doubt in her mission arise from wondering what Nina would think of it, or whether she would want Cassie to move on with her life—something Cassie cannot do, precisely because of her love for Nina. Cassie's revenge is an act of what Jean Harvey calls "moral solidarity" (2007) with Nina, a "caring attentiveness" to the nuances of Nina's experiences and a determination to make things right as

best she can *for* Nina's sake and the sake of their friendship, even if she does not ultimately defer to what she imagines Nina would have wanted (30).

There is no question that Cassie takes revenge against her targets and that in doing so she intentionally causes them harm. Further, it is equally clear that Cassie is deeply emotionally invested in her project of vengeance, and that her grief and anger fuel and shape her plans for revenge. She is, by any measure, a recognisable revenge taker, and one who displays many of the traits described by philosophers critical of revenge: her quest for revenge consumes her to the point of risking her other projects, relationships, and life, and it causes her to act deceitfully and even manipulatively, coercing others into confronting their moral responsibilities. Yet Cassie is far from the embittered, obsessive, reactive avenger that philosophers and others have feared. She chooses actions that aim at accountability rather than suffering, considers both proportionality and fit, and alters her intended actions in the face of remorse and contrition. Her capacity to enact revenge thoughtfully and morally seems *tied* to her care and concern both for justice, as a general principle, and to her friend Nina and their friendship, as a concrete relationship. Cassie has a deeply relational approach to revenge, one that connects her both to her targets and her cause in morally sensitive ways; the vision of revenge she enacts is relational, responsive, and—I would argue—profoundly feminist.

Conclusion

The purpose of this paper has been, first, to challenge the idea that revenge is always bad—that calling something revenge is tantamount to condemning it. I have done so first by challenging philosophical arguments against revenge, then by offering a philosophical framework that opens up the possibility of a moral role for revenge, both as a form of moral address and as something that can be enacted virtuously, by the right agent acting in the right way. But, in shifting to depictions of revenge that illustrate some of the features of righteous revenge, I have tried to go further than simply allowing for the conceptual possibility of revenge done (morally) well. Rather, I have argued, such narratives explore what appropriately constrained acts of revenge might actually *contribute* to processes of moral repair, following wrongdoing.

Both Cassie and Hayley illustrate how interpersonal revenge can draw out accountability and express moral protest in the absence of institutional justice. In Hayley's case, the accountability and protest in question might well make us withdraw from endorsing the possibility of legitimate interpersonal revenge. And yet, I've argued, the differences between Cassie and Hayley point to what is missing both in Hayley's actions and in the account of virtuous vengeance French has offered: not only does reparative revenge aim at accountability for wrongdoers and others who hold responsibility, it does so in solidarity with victims, expressing care and concern for their value, while remaining accountable (if not deferential) to their perspectives and wishes. Moreover, Cassie's revenge—what I have described as *reparative* revenge—aims to disturb and disrupt rather than harm, putting her targets in a position to reconsider their actions and agency. These two characteristics—victim solidarity and reparative accountability—provide a useful framework for reflecting on the value not only of acts of revenge, but also of other interpersonal interventions into the aftermath of wrongdoing.

Cassie's project of vengeance is precarious and risky: it damages her own prospects, her other relationships, her safety, and her life. In this way, she represents another cautionary tale: the serious costs of enacting individual responses to a collective and systemic problem (in this case, the failure to hold perpetrators of sexual violence responsible). It's not hard to imagine how much better things might have gone for Cassie if she had friends,

allies, or even a movement behind her, and how much *less* she would have had to personally risk to demand a moral conversation that needed to happen with institutions and individuals who had up until now refused to listen. Indeed, in such a situation, her actions might well have come to resemble the demands of #MeToo.

The horizon for my discussion has been the legitimacy of *personal* responses to wrongdoing in the absence of institutional justice, specifically as they arise in relation to sexual violence and assault. My hope is that taking seriously even the drastic, shocking methods employed by heroines in genre films gives us a new perspective on the more moderate forms of accountability that survivors and others in solidarity have demanded, through #MeToo activism and elsewhere. Not only can #MeToo not be reduced to "mere revenge" as a response to decades of sexual violence and harassment but, if it could, that is no reason not to take it seriously.

References

Améry, Jean. 2009. "Resentments." In *At the Mind's Limits: Contemplations by a Survivor on Auschwitz and Its Realities*, 62–81. Bloomington, IN: Indiana University Press.

Arendt, Hannah. (1958) 1998. *The Human Condition*, 2nd ed. Chicago, IL: University of Chicago Press.

Bacon, Francis. (1625) 1986. "Of Revenge" in *Francis Bacon: The Essays*. Penguin Classics.

Bell, Macalester. 2013. *Hard Feelings: The Moral Psychology of Contempt*. Oxford: Oxford University Press.

Bernardt, Fabian. 2020. "Forgiveness and Revenge." In *The Routledge Handbook of Phenomenology of Emotion*, edited by Thomas Szanto and Hilge Landweer, 497–508. London: Routledge.

Branco, Nelson. 2018. "Revenge Accusations: Has #MeToo Movement Led to a Witch-Hunt?" *Toronto Sun*, February 6, 2018. https://torontosun.com/news/world/revenge-accussations-has-metoo-movement-led-to-a-witch-hunt.

Cherry, Myisha. 2021. *The Case for Rage*. Oxford: Oxford University Press.

———. 2023. "Feeling Vengeful." *Passion: The Journal of the European Philosophical Society for the Study of Emotions* 1(1).

Chester, David S. and C. Nathan DeWall. 2017. "Combating the Sting of Rejection with the Pleasure of Revenge: A New Look at How Emotion Shapes Aggression." *Journal of Personality and Social Psychology* 112 (3): 413–30.

Elster, Jon. 1990. "Norms of Revenge." *Ethics* 100 (4): 862–85.

French, Peter. 2001. *The Virtues of Vengeance*. Lawrence, KS: University Press of Kansas.

Hall, Ciara. 2018. "Is the #MeToo Movement Becoming an Excuse for Vigilante Justice?" *Elephant Journal*, January 25, 2018. https://www.elephantjournal.com/2018/01/is-the-metoo-movement-becoming-an-excuse-for-vigilante-justice/.

Harris, Aisha. 2021. "The Agony and Subversion of the 'Promising Young Woman' Ending." *NPR Pop Culture Happy Hour* blog, January 21, 2021. https://www.npr.org/2021/01/21/958526533/the-agony-and-subversion-of-the-promising-young-woman-ending.

Govier, Trudy. 2002. *Forgiveness and Revenge*. Abingdon: Routledge.

Harbin, Ami. 2016. *Disorientation and Moral Life*. Oxford: Oxford University Press.

Harvey, Jean. 2007. "Moral Solidarity and Empathetic Understanding: The Value and Scope of the Relationship." *Journal of Social Philosophy* 38 (1): 22–37.

Levick, Richard. 2019. "#MeToo is Too Important to let devolve into 'Vigilantism.'" *Forbes*, March 25, 2019. https://www.forbes.com/sites/richardlevick/2019/03/25/metoo-is-too-important-to-let-devolve-into-vigilantism/?sh=47aad100d2ac.

Livsey, Anna. 2018. "Liam Neeson says Harassment Allegations are Now 'a Witch Hunt.'" *The Guardian*, January 13, 2018. https://www.theguardian.com/film/2018/jan/13/liam-neeson-says-harassment-allegations-have-become-a-witch-hunt.

MacLachlan, Alice. 2016. "'Hello. My Name is Inigo Montoya': Revenge as Moral Address." In *Reasonable Responses: The Thought of Trudy Govier*, edited by Catherine Hundleby, 129–48. Windsor, ON: Windsor Studies in Argumentation.

———. 2017. "In Defense of Third-Party Forgiveness." In *The Moral Psychology of Forgiveness*, edited by Kathryn Norlock, 135–60. Washington, DC: Rowman and Littlefield.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Belknap Press.

Pettigrove, Glen. 2009. "The Standing to Forgive." *The Monist* 92 (4): 583–603.

Scharf, Zack. 2018. "Terry Gilliam Fears #MeToo is Turning into 'Mob Rule,' but Admits Harvey Weinstein Is an 'Asshole.'" *IndieWire*, March 16, 2018. https://www.indiewire.com/2018/03/terry-gilliam-harvey-weinstein-metoo-mob-rule-1201940274/.

Schumann, Karin and Michael Ross. 2010. "The Benefits, Costs, and Paradox of Revenge." *Social and Personality Psychology Compass* 4 (12): 1193–205.

Smith, Adam. (1759) 1982. *The Theory of Moral Sentiments*, edited by D. D. Raphael and A. L. Macfie. Indianapolis, IN: Liberty Fund.

Solomon, Robert. 1999. *The Joy of Philosophy: Thinking Thin versus The Passionate Life*. Oxford: Oxford University Press.

Stainton, Robert. 2006. "Revenge." *Crítica: Revista Hispanoamericana de Filosofía* 38 (112): 3–20.

Swift, Taylor. 2022. "Vigilante Shit." On *Midnights*, Republic Records.

Uniacke, Suzanne. 2000. "Why Is Revenge Wrong?" *Journal of Value Inquiry* 34 (1): 61–69.

Von Maur, Imke. 2022. "The Epistemic Value of Affective Disruptability." *Topoi* 41 (5): 859–69.

Walker, Margaret. 2006. *Moral Repair*. Cambridge: Cambridge University Press.

Wikipedia. 2022. "International Response to the MeToo Movement." https://en.wikipedia.org/wiki/International_response_to_the_MeToo_movement. Accessed October 15 2022.

Winterson, Jeanette. 1997. "The Art of Fiction No. 150." Interview by Audrey Bilger. *The Paris Review* 145 (Winter). https://www.theparisreview.org/interviews/1188/the-art-of-fiction-no-150-jeanette-winterson.

Received: 15 August 2022
Accepted: 11 January 2023
Published: 15 June 2023

Feeling Revengeful

Myisha Cherry - University of California, Riverside

Abstract

I provide an account of feeling revengeful and I do so while rejecting three revenge myths: (a) feeling revengeful means feeling anger, (b) feeling anger means feeling revengeful, and (c) succeeding at revenge means achieving something negative like putting others down. My aim is to articulate the complex dimensions of feeling revengeful which will also disprove the above views. I also make a case for precision in the ways we describe our emotional and affective states and trouble the tendency to necessarily link anger to revenge, anger's action tendency to vengeance, and view angry people as "the avengers."

Keywords: revenge, anger, retribution, retributive emotions

Introduction

I want to think about revenge as a complex feeling and do so while challenging three revenge myths: (a) feeling revengeful means feeling anger, (b) feeling anger means feeling revengeful, and (c) succeeding at revenge means achieving something negative such as putting others down. My aim is to articulate the complex dimensions of revenge as feeling(s); this will disprove (a) and (b). I will do this by showing the diverse psychological and emotive states involved in feeling revengeful—in which anger *may* play a part, but isn't required—and by claiming that the "approach tendency" of anger is a tendency with various approach options, in which, again, revenge *may* play a part, but isn't required. My first aim is to make a case for precision in the ways we describe our affective states and to trouble the tendency to necessarily link anger to revenge. My second aim is to provide, albeit briefly, an example of oppressed communities feeling revengeful in ways that escape standard moral objections to revenge—a case of what I refer to as "the success model" of feeling revengeful.

I'll begin by providing an account of feeling revengeful. In doing so I'll also disprove the idea that feeling revengeful means feeling anger. I'll then argue that there are different types of anger—all with an approach tendency but with different approach options and aims—to disprove the claim that feeling anger means feeling revengeful. I'll conclude by claiming that we not only have linguistic reasons to improve the ways we think and talk about anger and revenge in general and "feeling revengeful" in particular, but moral reasons as well. I'll end by drawing on the work of W. E. B. Du Bois and James Baldwin to illustrate how oppressed communities often feel revengeful in ways that escape moral objections found in the literature.

The Complex Dimensions of Feeling Revenge

Revenge is not an unfamiliar or rare phenomenon. It's a popular theme in literature and film perhaps because it is something we all can relate to—given the persistence of wrongdoing and the agency and desires of victims and bystanders. Homer's *Iliad* is rich with examples of revenge. The Trojan War begins as an act of revenge against Paris for stealing Helen, the wife of Menelaus. Later, Achilles kills Hector for killing his childhood friend Patroclus. Shakespeare's *Hamlet* begins by the protagonist's father's ghost commanding him to "Revenge his foul and most unnatural murder" (1.5.25). And he succeeds, but soon after he is stabbed by Laertes's blade. Men are not the only fictional characters who enact revenge in literature. Stephen King's *Carrie* is a story about a woman who takes revenge on those who have bullied and hurt her. The book ends with a climactic scene of Carrie, with the use of telekinetic powers, engaging in a bloody act of revenge on her classmates. Revenge is also a major theme in film. In Jack Hill's 1974 American blaxploitation film *Foxy Brown*, the main character, played by Pam Grier, goes after a gang who murdered her boyfriend. In a memorable scene, Foxy says to a group of young revolutionaries, "You just take care of the justice, I'll handle the revenge myself." And she does! Here Foxy makes an important distinction and has no problem declaring which one she prefers. In The Bullitts's 2021 Western *The Harder They Fall*, a stellar Black cast of characters depict two rival gangs headed by outlaws Nate Love and Rufus Buck respectively. Nate seeks revenge against Rufus for killing his parents when he was a child. Pointing to the prudential necessity of enacting revenge, Nate's love interest reminds him that "as long as that man [Rufus] draw breath, your spirit is gonna be cursed and wild as it ever was." Nate gets his revenge in the end but only after Rufus reminds him that "letting [him] live, that was the true revenge." All this revenge, albeit fictional, is sensational to witness. It also speaks to the human potential and hidden desires of many to enact their own revenge. Perhaps that's why revenge remains a recurring theme in popular culture. We know what revenge looks like in action. But what does it feel like? That is to say, what are its affective features? This is the task I take up in this essay.

To be sure, philosophers and social psychologists have linked feeling and revenge. The most common accounts have been "feelings of revenge," "retributive emotions," and "emotions of revenge." These thinkers describe feelings of revenge as emotions that come about through humiliation, last longer than other emotions, and whose goal is to make the target suffer (Elshout et al., 2015). Feeling and revenge are also linked together in discussions about retributive emotions. Philosophers define these emotions as those that motivate and/or comes with a retributive wish. Emotions that are often listed as retributive are anger and hatred. Jeffrie Murphy, for example, in describing "retributive hatred," writes:

> Jones has injured me, has taken unfair advantage of me, has brought me low, and is himself unrepentant and flourishing. I hate him and want him brought low . . . I want Jones to be hurt . . . [P]art of the basis for desiring the hurt is the desire to restore what seems (at least to me) to be the proper moral balance of whatever goods are in question. (1988, 89)

What for Murphy makes hatred a retributive feeling is its desire for the target to suffer in order to restore "proper moral balance."

Philosophers have also linked *the act* of revenge to emotions and affective attitudes. In *Forgiveness and Revenge* (2002), Trudy Govier highlights the various dimensions of revenge. In her view, revenge entails intent: revenge is "deliberately bringing harm . . . for the sake of enjoying having brought that harm" (12). However, revenge is not limited to motive. There are emotions that are also part of the picture. For Govier, the emotion of revenge

is "hatred that goes so far as to include joy at the evil meted out to another person . . . positive joy in the fact that we have caused the suffering of another person" (13). When revenge is practiced there are core elements to its success. The three core elements to successful revenge are: (1) intention to harm another in response to their initial harm, (2) the revenger's agency in that harm, and (3) the actual suffering of the one who is harmed.[1] What makes revenge harmful and therefore morally objectionable for Govier is that it treats another as the instrument of our satisfaction and "morally diminishes the revenging victim" (13).

In an attempt to broaden Govier's claims, Alice MacLachlan (2016) notes that the intent aspect of revenge is more robust. It's not enough that we intend to harm our target: additionally, "the revenger needs her target to know that this is an act of revenge" (139). While there are reasons to reject revenge as a moral good, MacLachlan doesn't think that in revenge we always treat our targets as instruments of our satisfaction. Rather, she argues that "revenge is . . . a communicative action [where the recipient is a] moral addressee, rather than an instrument to my purposes" (137). In this way, "the target is not merely instrument, but audience" (140). While she points out that "I necessarily take their personhood seriously insofar as I see them as someone capable of persuasion . . . [a strike against revenge is that] it is intended to end, rather than continue moral conversation" (143). Therefore, when she condemns revenge, she acknowledges that "Govier is right that revenge may morally diminish the revenger—not because it requires a particular range of distasteful psychological states, but because it effectively freezes her in a particular role and that role's subsequent perspective" (147).

I find these thinkers' views on emotions, attitudes, and revenge illuminating. However, their accounts are insufficient for describing feeling revengeful. For example, their accounts seem to be limited to retributive emotions like anger and hatred as if they are sufficient for describing what it means to feel revengeful. Their accounts of revenge also rely too heavily on what these emotions do, their target, and their length, while ignoring other affective attitudes and other feelings that are experienced (when these emotions are absent). Although there is focus on intent and desire in some of these accounts, they fall short in capturing the complex dimensions of what it is to feel revengeful—beyond emotions and desires. Also, several of the thinkers' primary concerns are not my focus in this essay. For example, contra Govier and MacLachlan, I am not concerned with successful "acts of revenge" but with feeling revenge (having the feeling even if we don't engage in revengeful actions). While there is mention of "emotion[s] of revenge" (e.g., hatred on the part of Govier and Murphy) and claims that revenge is a "complex emotion," I'm interested in accounting for more affective aspects of revenge. It includes emotion, but also other psychological states and affective attitudes. The word "feeling" captures the broadness that I'm after.

Before I describe "feeling revengeful," it's important to highlight what I am not referring to. By "feeling revengeful," I am not referring to the emotions of revenge like hatred, as if we can reduce feeling revengeful to them. My view is that they cannot. Rather, I'm describing a more complex phenomenon that consists of desires and emotions, as well as other affective states (such as pleasure and so forth) which may have different aims, focus, and physiological responses. As a result, the experience of feeling revengeful will vary by persons and according to circumstances.

I'm going to introduce five features of what it is to paradigmatically feel revengeful, arguing that in these cases one feature will typically be present *and* accompanied by one or more of the others. My aim through this "combination approach" is to present a more varied and complete picture of the phenomenon without reducing feeling revengeful to the experience of retributive emotions.

1 As summarised by MacLachlan (2016).

The first feature of feeling revengeful is *vindictive plans*. These plans include the intentions and motives of the agent. In feeling revengeful, the agent has the intention to harm another or a desire that they be harmed in response to their intentional harm. But as MacLachlan points out, the target of my revenge must know the reason for the harm. Using the example of coworkers, she notes:

> Take two colleagues: Leah and Mateo. Mateo has ruthlessly and persistently bullied Leah over many years . . . Leah sits on a committee where she suddenly found herself in a position to vote to defund a pet project of Mateo's—and she did so. Mateo was not a member of this committee, and will never know the individual votes. Leah's swing vote determined the outcome . . . But unless Mateo knows that she, Leah, voted, and how and why she voted, and unless he knows Leah's vote was the deciding factor, Mateo's relationship to Leah will not change. Leah hasn't yet "gotten her own" back, or sent the message of "agency, wrong, responsibility, and rightful suffering" that she wishes to communicate. (2016, 138–39)

It's not enough that I have the desire or intention to harm, I must desire that the target know that I am the reason why they have been harmed. Such harm may be physical or psychological. The suffering might entail the experience of excessive guilt but also goal frustration.

The second feature is *vindictive powers*. In feeling revengeful, I not only desire that the target is harmed: I also take such harm to be instrumental, a means to an end. The end is to achieve a restoration of power in the sense of balancing the universe or achieving individual moral power. On this view, the harm is desired in order to right a wrong, to balance the score, or to restore a sense of fairness and reciprocity. In Murphy's terms, it's to "restore what seems (at least to me) to be the proper moral balance of whatever goods are in question" (Murphy 1988, 89). The agent believes and therefore wishes that the target is harmed so that a previous harm is undone. If injuries are seen as diminishments, then relative status has been harmed via the wrongdoing. In feeling revengeful, I desire my target suffers so that revenge can "atone for or annul the damage" (Nussbaum 2016, 26).

The harm can also be desired to gain recognition; this is to say that it aims for the target to acknowledge wrongdoing.[2] This is quite different from MacLachlan's claim that the wrongdoer know that we are behind the revenge. I can desire that they know I orchestrated their destruction, without desiring that they confess to the offence or acknowledge the harm they've caused. The avenger can feel revengeful by desiring to say to his target—as is the case in the movie *The Princess Bride*—"Hello. My Name is Inigo Montoya. You killed my father. Prepare to die." But he can also feel revengeful by doing as the fictional detective Adrian Monk did after he found the killer of his wife Trudy after twelve years of searching: get recognition. While pointing a gun at his target Monk instructs him to dig. Unbeknownst to the police witnessing, Monk knows there is a body buried in the front yard. By instructing his target to dig, he's getting him to admit to the killings of not only his wife but a witness. The killer immediately begins to confess and explain why he murdered them. This was the recognition Monk was after.

But feeling revengeful doesn't just consist of motives and aims. Feeling revengeful is typically accompanied by emotions—what Murphy describes as *vindictive passions*. These passions, on his view, are "often felt by victims towards those who have wronged them" (2003, 17). Vindictive passions include hatred, resentment, and moral disgust.[3] For Govier, the emotion of revenge is "hatred that goes so far as to include joy at the evil meted out to

2 Silva (2021) claims that this is what anger is after, but I think revenge, even absent anger, is after such acknowledgement.

3 Here I depart from Jeffrie Murphy's account of vindictive emotions. For him, they are limited to anger, resentment, and desire for revenge. My account is more expansive.

another person" (13). According to Adam Smith, "resentment would prompt us to desire, not only that he should be punished, but that he should be punished by our means, and upon account of that particular injury which he had done to us" ([1759] 1976, 63). Moral disgust is often not listed as a vindictive passion. This is because it is associated with avoidance behaviour. The victim is less likely to directly confront the offender when they feel disgust. However, as Curtis and Biran (2001) note, this just makes it likely that the person who experiences moral disgust would recruit revenge by others. In addition, they claim that any avoidance associated with moral disgust functions "to punish and ostracize moral offenders" (29). So, while moral disgust is not typically considered a vindictive passion, we can see why it is an emotion one can feel when they feel revengeful.

However, I do not think the presence of these emotions fully explains what it is to feel revengeful. Unless these emotions conceptually motivate vengeful action or necessarily entail a desire for revenge, vindicative passions alone can't give the full account of the feeling-revengeful phenomenon. This list is also not exhaustive. For example, we can imagine a hybrid of emotions that might fit under vindictive passions. They might include humiliation and resentment, or disgust and grief. Or they might include what David Shoemaker describes as "hurt feelings." Moreover, emotions that might not traditionally count as vindictive might count when in combination with other emotions. These emotions might include shame, a blind love for justice, or pride. Murphy claims that emotions like anger and hatred "often prompt acts of vengeance or revenge, but one can have the passions without acting on them—just as one can feel sexual lust without acting on it" (2003, 16). However, feeling revengeful should not be reduced to the experience of these vindictive passions. In feeling revengeful, there are other things we typically feel and desire beyond and in addition to these emotions. Furthermore, MacLachlan points out that since there are various forms of revenge one could intend (e.g., nonviolent, proportionate, etc.), there are also different categorical feelings we can feel as a result. While vindictive passions such as hatred, disgust, and resentment may be present when desiring violent cases of revenge, they may be absent when desiring nonaggressive instances of revenge, cases like "the success model" that I defend in section 5.4 Therefore, I do not think the retributive emotions of anger and hatred can fully explain what it is to feel revengeful. There are additional emotions—beyond the standard, retributive ones—that also fall under the category of vindictive passions.

There are also other things we can feel or desire to feel when feeling revengeful. These other affects may include pleasures. The fourth feature in feeling revengeful is *vindictive pleasures*. These pleasures come about through the contemplation of and/or successful completion of revenge (via the suffering of the wrongdoer). Aristotle associated revenge (via anger) with pleasure. He writes: "[Anger] must always be attended by a certain pleasure—that which arises from the expectation of revenge" (*Rhetoric*, II.2). These pleasures include enjoyment and satisfaction. I can get joy from planning a revengeful scenario. This pleasure feature is explained in *The Count of Monte Cristo* as Dantès talks about his revenge: "How did I plan this moment? With pleasure." I might also get pleasure in the form of a sense of satisfaction at a job well done performed in the name of moral duty and self-respect (Barton 1999). I might also experience joy after I see my target harmed. Govier suggests that hatred can include such joy. I agree. But I also think that an agent can experience such enjoyment without the hatred. Consider the case of revenge fantasies. Recent research in psychology suggest that revenge fantasies can calm negative feelings of frustration and humiliation and bring with them pleasure at imagining the suffering of wrongdoers. Without the experience of hatred by agents imagining revenge, they can experience calm and comfort.[5] When a person aims for or achieves such pleasure by thinking about, planning, or engaging in revenge, they are feeling revengeful.

4 See MacLachlan (2016, 136).

5 See Craig Haen and Anna Marie Weber (2009).

Lastly are *vindictive products*. Contrary to the notion that we tend towards vengeance in order to satisfy our vindictive passions, we can also feel vengeful to satisfy other things whether we intend them or not. I refer to these feelings and desires as vindictive products. They include the desire to achieve, or the experience of, relief, escape, closure, or catharsis from revenge (MacLachlan 2016). While feelings of relief or closure are not vindictive, they can be described as such when vindictiveness becomes a means to achieve them. MacLachlan thinks, contra Govier, that "not everyone commits revenge out of enthusiasm . . . [Therefore,] to equate the satisfaction of desire with pleasure at that satisfaction is mistaken, and risks committing to a simplistic psychological hedonism" (2016, 136). It's here where the distinction between pleasure and products matter. Relief and closure can be a product of the target's suffering even when pleasure at their suffering is absent.

Thus, feeling revengeful paradigmatically consists of the following features:

A) Vindictive Plans
The *intention* to harm another or a *desire* that they be harmed (by myself or others) in response to their intentional harm, and for them to know the reason for the harm. The nature of this harm or suffering can vary from physical harm to goal frustration.

B) Vindictive Powers
Harm *desired* as instrumental to righting wrongs or gaining recognition.

C) Vindictive Passions
Emotions like hatred, resentment, moral disgust, hurt feelings, or a combination of emotions like pride and humiliation.

D) Vindictive Pleasures
The contemplation and/or successful completion of revenge (via the suffering of the wrongdoer) that comes with *enjoyment* and *satisfaction*.

E) Vindictive Products
Desires to and/or achieve *relief*, *escape*, *closure*, or *catharsis* from revenge.

To feel revengeful is to paradigmatically experience A, B, D, or E. When C is experienced, it is typically accompanied by A, B, D, or E. This is because feeling revengeful is more complex than vindictive passions alone and therefore should not be reduced to them. While this account describes what it is to feel revengeful, it also supports two important claims. It shows that revenge is not dependent on anger for its existence. If we omit anger (one of the vindictive passions), we can still feel revengeful since (C) doesn't fully explain the phenomenon. But also, anger is not the only vindictive passion. One can experience hatred or disgust. And when one does, in combination with other features, one feels revengeful. This shows that revenge and anger are not necessarily linked. Therefore, the claim that feeling revengeful implies feeling anger is false.

I have argued that you can feel resentment and still not feel revengeful. In this way, I agree with Stephen Darwall that "reactive attitudes like resentment and indignation can be distinguished from the desire to retaliate or gain vengeance" (2006, 83). As a result, there are additional desires, affects, and emotions at play that account for typical experiences of the phenomenon—such as a desire to right a wrong or get closure—for the passion to feature in feeling revengeful. Resentment doesn't necessarily entail the desire, therefore: other intentions and motives must accompany resentment. In their absence, to feel resentment is not necessarily to feel revengeful. As we will see in the next section, this claim also depends on the type of anger experienced.

Action Tendencies and Breaking the Link

There are philosophers and psychologists who have not only linked anger with revenge but have defined anger in relationship to its vengeful tendency. Consider the following claims:

> It appears to me that you are right in feeling especial fear of this passion, which is above all others hideous and wild . . . [It] consists wholly in action and the impulse of grief, raging with an utterly inhuman lust for arms, blood and tortures, careless of itself provided it hurts another, rushing upon the very point of the sword, and greedy for revenge even when it drags the avenger to ruin with itself. (Seneca [45] 2010, 1.1)

On Seneca's account, what it is to be angry is to be greedy for revenge. And such revenge is bound to cause the vengeful agent to suffer too. Therefore, we have reasons to be afraid of anger and to resist it in our lives.

Martha Nussbaum has a similar view. She claims that anger involves this vengeful tendency and we should therefore try our best to transition out of anger into love and generosity. She writes:

> The idea of payback or retribution—in some form, however subtle—is a conceptual part of anger. I . . . argue the payback idea is normatively problematic, and anger, therefore, with it . . . The claim is not that anger conceptually involves a wish for violent revenge; nor is it that anger involves the wish to inflict suffering oneself upon the offender. For I may not want to get involved in revenge myself: I may want someone else, or the law, or life itself, to do it for me. I just want the doer to suffer . . . that anger involves, conceptually, a wish for things to go badly, somehow, for the offender, in a way that is envisaged, somehow, however vaguely, as a payback for the offense. They get what they deserve. (Nussbaum 2016, 22–23)

There are social psychologists who have also made similar claims. Carroll E. Izard (1997) has argued that the action tendency of anger is retaliation. And more recently, Jonathan Haidt has argued that "anger comes with an inclination to attack, humiliate, or otherwise get the person back who is perceived as acting unfairly or immorally" (2003, 856). Some have made weaker claims, choosing instead to reject the conceptual point and opt for an account based on occurrence. For example, Charles Barton (1999) claims that "revenge is . . . *typically* accompanied and fueled by feelings of indignation, anger, and resentment for wrongs suffered" (1999, 86). If these thinkers are right, then we cannot separate anger from feeling revengeful; for what it is to feel anger is to feel revengeful. However, in what follows I'll show why I reject this picture.

In *The Case for Rage* (2021), I argue that, contrary to the broad strokes view of anger, anger is not one thing. There are different types. We can identify them by looking at the target (at whom and what it is directed); the perspective that informs the anger (the attitude from which the anger arises); its action tendency (or propensity to act); and its aim (i.e., the purpose and plan from which the anger arises). The action tendency of all anger types is "approach motivation," in which an agent is motivated to approach the target of her anger. But the kind of approach one desires or engages in is depended on many factors. It depends on the type of anger one has, given its features. To help illustrate this, let's look at the differences between two anger types as argued in the book.

One type of anger that arises in the context of political injustice is *wipe rage*. The target is racial "others," those who are of a different race than the agent. The perspective that informs wipe rage is a zero-sum-game way of thinking. The agent falsely thinks that they are experiencing injustice because of Mexicans or Blacks, and part of their anger is informed by the attitude that if these racial others make progress or succeed, then this will ensure his and his racial group's failure. These two features will influence the type of action one would have a tendency to engage in, as well as their aims. The action tendency in wipe rage is to eliminate the racial other, and its aim is the racial other's social or physical exclusion.

This contrasts in stark ways with a more positive type of anger: *Lordean rage*. This kind of anger also arises in the context of political injustice. However, rather than being targeted at racial others, its target is those who are complicit in or perpetrators of racism. The perspective that informs the anger is an inclusive perspective, for the person with Lordean rage does not believe they are free until everyone is free. These features affect the action tendency and aim of the angry agent. The action tendency is for the agent to fight the problem by changing it. Thus, the aim of Lordean rage is to change laws, policies, culture, those in power, etc. Remember, the approach tendency is connected to the aim, which is informed by the other features. If one has wipe rage, one is more prone to engage in revengeful behaviour given one's target and perspective. However, those who have Lordean rage are not likely to engage in standard destructive cases of revenge, but rather will tend, due to the anger's target and perspective, to engage in constructive action. So, while revenge may be part of the picture of some types of anger (wipe rage), it is not necessarily part of all types.

This shows that if we omit revenge as one of the approach options, we will still have anger (e.g., Lordean rage). We can therefore conclude that anger and revenge are not necessarily linked. This also makes the colloquial phrase "don't get mad, get even" coherent—for it's possible to feel revengeful without anger. Thus, we can also conclude that the claim "feeling anger means feeling revengeful" is false.

The "Moral" of the Story

My account of feeling revengeful is important because it allows us to achieve more precision in the ways we describe certain phenomena. This precision has both linguistic and moral import.

As we've discovered, there is no need to overemphasise the connection between anger and revenge as if they are necessarily and always linked. It's best we choose our words carefully! We should be resistant to the description: "She is angry and therefore feels revengeful." I have shown that this is an inaccurate moral picture. I've claimed that since there are a variety of angers with different aims and action tendencies, anger doesn't necessarily entail the desire for revenge. In addition, the claim "she feels revengeful and is therefore angry" is also inaccurate since I've argued that feeling revengeful doesn't necessitate feeling angry because there are other vindictive emotions like hatred and hurt feelings that one can feel instead. The assumptions behind these two claims are also incomplete. Feeling angry is not sufficient for feeling revengeful. Vindictive powers or products typically accompanies the anger.

There are more precise alternatives on the table to describe an agent in relationship to anger and revenge. They are the following:

1. She is angry but doesn't feel revengeful.
2. She feels revengeful but doesn't feel angry.
3. She is angry and feels revengeful.

Any of these options paints a more accurate moral picture of the agent. We must remember that (3) is not the only possibility. However, when we describe it as such, we have a fuller account of how she feels than we do on the incorrect picture.

I'm not just concerned with conceptual thinking and linguistic accuracy but moral action. Silencing, vigilantism, and political backlash can occur if we overemphasise the connection between anger and revenge (i.e., "She is angry and therefore feels revengeful").

For example, instead of listening to and giving uptake to an angry victim's claims of wrongdoing and injustice, we are likely to silence them with the thought that silencing their anger is silencing their revenge. This can have horrible consequences for the moral and political spheres. It can dissuade protest, prevent recognition, and delimit democracy.

Vigilantism can also result from the inaccurate picture. Seeing an angry person may justify irrational fear and the need for self-defence. If angry people are avengers in waiting, we may think that we must attack them before they attack us. This will sow seeds of distrust and violence. There can be no justice, respect, or peace when such thinking guides our behaviour. But such actions are not limited to individual citizens. There can also be state-level political backlash against angry citizens. Since the thought is that angry citizens are also feeling revengeful, political steps might be taken to respond, albeit irrationally, to angry citizens in order to curb their potential vengeful behaviour. This may manifest in increased policing, oppressive policies, and state violence. Since emotions are gendered and racialised—that is to say, we read and assess emotions differently based on a person's race and gender—this is likely to have a disproportionate impact on women and people of colour. There are certain emotions, such as anger, that we are encultured to think are virtuous for men or white people to have and vicious for women and oppressed minorities to have. This is likely to affect *who* we see as angry avengers in waiting. Angry Black protestors, angry feminists, and angry Palestinians, for example, are more likely to be victims of silencing, vigilantism, and political backlash.

Given these practices, we have linguistic and moral reasons to improve the ways we think about anger and revenge in general, and about "feeling revengeful" in particular.

The Success Model

One of the concerns that white Americans held after the abolition of slavery was fear that the formerly enslaved would take up arms and enact violent revenge. It's not surprising that given these worries, many race men such as Booker T. Washington sought to convince whites that Blacks were not resentful. Unfortunately, these worries still exist today. I want to end by claiming, controversially, that some Blacks do indeed feel revengeful in response to historical and modern racism and racial oppression. But the nature of the feeling runs counter to the popular white imaginary.

I refer to this feeling as the "success model" of feeling revengeful because part of the feeling entails the idea that "success is the best form of revenge." This model of feeling revengeful has aided in Blacks' resistance and survival, and escapes some of the criticisms posed by MacLachlan and Govier.

Two examples that capture the success model comes from W. E. B. Du Bois and James Baldwin. In the first chapter of *The Souls of Black Folk* (1903), Du Bois describes his discovery of being different from his white classmates. This discovery then leads to a desire that, I believe, captures feeling revengeful.

He writes:

> I was different from the others; or like, mayhap in heart and life and longing, but shut out from their world by a vast veil to creep through; I held all beyond it in common contempt, and lived above it in a region of blue sky and great wandering shadows. That sky was bluest when I could beat my mates at examination-time, or beat them at a foot-race, or even beat their stringy heads. Alas, with the years all this fine contempt began to fade; for the words I longed for, and all their dazzling opportunities, were theirs, not mine. But they should not keep these prizes, I said; some, all, I would wrest from them. Just how I would do it I could never decide: by reading law, by healing the sick, by telling the wonderful tales that swam in my head,—some way . . . The history of the American Negro is the history of this strife . . . This, then, is the end of his striving: to be a co-worker in the kingdom of culture, to escape both death and isolation, to husband and use his best powers and his latent genius. (8–9)

There are vindictive plans. As a result of being shut out by the veil, Du Bois decides to beat his classmates at everything and to wrest opportunities from them through reading law and telling tales. There are also vindictive products. He does it so that he might escape death and isolation, and use his latent genius. There are vindictive powers, for such acts are instrumental to gaining recognition via being coworkers in the kingdom of culture. There is also vindictive pleasure. Du Bois admits that the sky was blue (i.e., that his days were wonderful) when his striving (or revenge) was successful. Lastly, there is a vindictive passion, a contempt that eventually fades. Du Bois refers to what I am calling feeling revengeful as spiritual strivings, and there's no doubt that is indeed that. Feeling revengeful in a context of oppression can produce a striving that consists in part of what I am calling the success model of feeling revengeful.

A passage from James Baldwin's "My Dungeon Shook: A Letter to My Nephew on the One Hundredth Anniversary of the Emancipation" ([1962] 1993) also provides an example of the success model. Baldwin's instructions to his nephew are different from Du Bois's description in that while Du Bois's feeling is connected to outward action; Baldwin's instructions are psychological. Yet, it is still an example of feeling revengeful. Baldwin writes:

> This innocent country set you down in a ghetto in which, in fact, it intended that you should perish . . . You were born where you were born and faced the future that you faced because you were black and *for no other reason*. The limits of your ambition were, thus, expected to be set forever. You were born into a society which spelled out with brutal clarity, and in as many ways as possible, that you were a worthless human being. You were not expected to aspire to excellence: you were expected to make peace with mediocrity. Wherever you have turned, James, in your short time on this earth, you have been told where you could go and what you could do (and *how* you could do it) and where you could live and whom you could marry. I know your countrymen do not agree with me about this, and I hear them saying, "You exaggerate." They do not know Harlem, and I do. So do you. Take no one's word for anything, including mine—but trust your experience. Know whence you came. If you know whence you came, there is really no limit to where you can go. The details and symbols of your life have been deliberately constructed to make you believe what white people say about you. Please try to remember that what they believe, as well as what they do and cause you to endure, does not testify to your inferiority but to their inhumanity and fear. (7–8, emphases in original)

There is a vindictive plan. Baldwin wants his nephew to quash white expectations, to frustrate their plans for him to perish. There is also a vindictive product. The aim of such quashing is so that he may escape and get relief from what his white countrymen say or intend for him. He does this by refusing to believe what they say about him. Note that there is no anger or hatred. Baldwin in the next few sentences tells his nephew to accept his countrymen with love rather than contempt or anger. This love in combination with pride or hurt feelings produces what I am calling vindictive passions.

Although there is a focus on vengeance and targets in the success model, this is not all of what striving, in general, aims at. In the lives of the oppressed, the success model of feeling revengeful runs alongside other feelings and plans such as love for oneself and joy despite oppressive conditions, as well as desires to survive, thrive, honour one's ancestors, and leave a legacy. In this way, feeling revengeful is part of the story of the oppressed, but is never the whole story.

The success model also escapes some of the worries posed by Govier. For example, rather than "treat another as instruments of our satisfaction" it instead treats evil intentions, desires, and actions as instruments. They become the motivation to survive despite obstacles. Since the aim and action of the oppressed are plan frustration/failure via success rather than harm proper, feeling revengeful doesn't "morally diminish the revenging victim." Instead, it can, as Baldwin claims, validate them. Although MacLachlan is worried that revenge ends rather than continues the moral conversation by making the target an audience, the success model operates differently. Some conversations, such as those in which one's dignity, humanity, or inclusion is under discussion, should not be continued. When this occurs, targets *should* be an audience to the oppressed party's survival and thriving—which, fortunately, occurs despite the targets' cruel actions or aims.

Conclusion

Feeling revengeful is more complex than simply experiencing anger. There are typically other emotions involved, as well as other psychological and affective features that are present; anger itself may not even be present. But feeling revengeful is not all about what we feel. Feeling revengeful can aid in one's survival against oppression and allow one to do so without perpetuating more oppression or suffering in the world. Getting clearer on these matters can help us understand our emotional lives better. It can also help curb further destructive behaviours, such as silencing and backlash, while encouraging more constructive types, such as striving.

References

Aristotle. *Rhetoric*. In *The Works of Aristotle*, edited by W. D. Ross. Oxford: Clarendon Press, 1931.

Baldwin, James. (1962) 1993. "My Dungeon Shook: A Letter to My Nephew on the One Hundredth Anniversary of the Emancipation." In *The Fire Next Time* ([1963] 1993). New York: Vintage International.

Barton, Charles. 1999. *Getting Even: Revenge as a Form of Justice*. Illinois: Open Court.

Cherry, Myisha. 2021. *The Case for Rage*. New York: Oxford University Press.

Curtis, Valerie and Adam Biran. 2001. "Dirt, Disgust, and Disease. Is Hygiene in Our Genes?" *Perspectives in Biology and Medicine* 44 (1): 17–31.

Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*, Cambridge, MA: Harvard University Press.

Du Bois, W. E. B. 2008. *The Souls of Black Folks*. New York: Oxford University Press.

Elshout, Maartja, Rob M. A. Nelissen, and Ilja van Beest. 2015. "Vengeance Is Self-Focused: Comparing Vengeful to Anger-Driven Responses." *Cognition and Emotion* 29 (7): 1239–55.

Govier, Trudy. 2002. *Forgiveness and Revenge*. New York: Routledge.

Haidt, Jonathan. 2003. "The Moral Emotions." In *Handbook of Affective Sciences*, edited by R. J. Davidson, D. R. Scherer, and H. H. Goldsmith, 852–70. Oxford: Oxford University Press.

Haen, Craig and Anna Marie Weber. "Beyond Retribution: Working through Revenge Fantasies with Traumatized Young People." *The Arts in Psychotherapy* 36 (2): 84–93.

Izzard, Caroll E. 1977. *Human Emotions*. New York: Plenum Press.

MacLachlan, Alice. 2016. "'Hello. My Name is Inigo Montoya': Revenge as Moral Address." In *Reasonable Responses: The Thought of Trudy Govier*, edited by Catherine E. Hundleby, 129–48. Windsor, ON: Windsor Studies in Argumentation.

Murphy, Jeffrie G. 1988. "Hatred: A Qualified Defence." In *Forgiveness and Mercy*, edited by Jeffrie G. Murphy and Jean Hampton, 88–110. Cambridge: Cambridge University Press.

Murphy, Jeffrie. 2003. *Getting Even: Forgiveness and Its Limits*. New York: Oxford University Press.

Nussbaum, Martha. 2016. *Anger and Forgiveness*. New York: Oxford University Press.

Seneca. (45) 2010. *On Anger*. In *Anger, Mercy, Revenge*, translated by Robert A. Kaster and Martha C. Nussbaum. Chicago: University of Chicago Press.

Silva, Laura Luz. 2021. "The Efficacy of Anger: Recognition and Retribution." In *The Politics of Emotional Shockwaves*, edited by Ana Falcato, 27–55. London: Palgrave Macmillan.

Smith, Adam. (1759) 1976. *The Theory of Moral Sentiments*. Oxford. Oxford University Press.

Received: 15 August 2022
Accepted: 11 January 2023
Published: 15 June 2023

Are *All* Mental Disorders Affective Disorders?

Michelle Maiese – Emmanuel College Boston, USA.

Abstract

A growing number of theorists have looked to the enactivist approach in philosophy of mind or the affordance-based approach from ecological psychology to make sense of a wide variety of phenomena; some theorists believe that these theoretical accounts can offer rich insights about the nature of mental disorders, their etiology, and their characteristic symptoms. I argue that theorists who adopt such approaches also should embrace the further claim that all mental disorders are affective disorders. First, enactivist accounts of mental disorder push us towards such a view insofar as they characterise such conditions in terms of disordered sense-making and conceptualise sense-making as fundamentally affective. Second, conceptions of mental disorders that emphasise affordance perception likewise motivate such a view insofar as they highlight the role that affectivity plays in the disclosure of action possibilities. What is more, both sense-making and affordance disclosure are best understood as processes of selective attention and responsiveness that rely heavily on affectivity. To illustrate and support these claims, I discuss how (a) language disturbances in schizophrenia and (b) "context blindness" in autism both result from disruptions to affectivity and selective attention that make it difficult for subjects to engage effectively with relevant affordances.

Keywords: affordance; autism; context integration; enactivism; language disturbance; schizophrenia; selective attention; sense-making

1. Introduction

In recent years, philosophical and phenomenological investigations of psychopathology have devoted more attention to ways in which affective dynamics may be involved in the etiology and symptomology of mental illness (Bortolan and Salice 2018). Still, there is a general tendency to suppose that some conditions (e.g. mood disorders) centre primarily around disruptions to affectivity, whereas others centre primarily around disruptions to cognition. But what if all mental "disorders" are affective, in an important etiological sense, and even symptoms involving thought, language, and executive control result (at least in part) from an affective disturbance? Do existing conceptualisations of mental disorder support such a view? A growing number of theorists have looked to the enactivist approach in philosophy of mind or the affordance-based approach from ecological psychology to make sense of a wide variety of phenomena; some theorists believe that these theoretical accounts can offer rich insights about the nature of disorders, their etiology, and their

characteristic symptoms. Here, I take it for granted that such approaches can help us to conceptualise various key aspects of mental disorder. My central claim is that theorists who adopt such approaches should embrace the further claim that all mental disorders are, in an important sense, affective disorders.

First, enactivist accounts of mental disorder push us towards such a view insofar as they characterise these conditions in terms of disordered sense-making, and conceptualise sense-making as fundamentally affective. Second, conceptions of mental disorders that emphasise affordance perception likewise motivate such a view insofar as they highlight the role that affectivity plays in the disclosure of action possibilities. What is more, both sense-making and affordance disclosure are best understood as processes of selective attention and responsiveness, i.e., processes that involve the prioritisation of some stimuli and the suppression of others. Crucially, such processes have an integral affective dimension; the way in which an agent selectively attends to and engages with their surroundings depends significantly on their affective orientation, which encompasses the range of ways in which they care about objects, events, states of affairs, other people, their own life, etc. This includes occurrent emotions such as anger or fear, more diffuse mood states, concerns, and existential orientations (Ratcliffe 2005). Some degree of selective attention is integral to all our skilful engagements with our surroundings, including sensory perception, memory, goal-formation, and action monitoring. Thus, disruptions to affectively guided selective attention processes are likely to contribute to notable cognitive and agential impairments. Indeed, even symptoms that may appear to be associated primarily with higher-order cognitive functions, such as language and context integration, are best understood as the direct result of an affective disturbance and associated disruptions to selective attention.

In the next section, I review enactivist accounts that conceptualise mental disorders in terms of disordered sense-making, and explain why proponents of such accounts also should accept the further claim that these conditions are affective disorders. In section 3, I adopt a similar argument strategy with respect to accounts that focus on affordance engagement. In section 4, I discuss how (a) language disturbances in schizophrenia and (b) so-called "context blindness" in autism both result from disruptions to affectivity and selective attention. These disruptions contribute to disordered sense-making and make it difficult for subjects to engage effectively with relevant affordances.

Before proceeding, it is worth noting that some theorists will reject the notion that autism counts as a genuine disorder, or perhaps reject talk of "disorder" altogether. Indeed, proponents of the so-called anti-psychiatry movement that began in 1960s have gone so far as to claim that "mental illness" is simply the accepted term for various behaviours and subjective experiences that are problematic or that do not fit the cultural norm (see, e.g., Szasz 1960). Along somewhat similar lines, contemporary proponents of neurodiversity have maintained that these conditions simply represent different modes of living that are not inherently pathological or disordered (see, e.g., Chapman 2020). I do not have sufficient space here to defend the claim that many of the conditions that have been labelled "disorders" involve very real incapacities, and that there is an objective fact of the matter about whether someone suffers from them (though see, e.g., Graham 2013). However, it is also important to highlight that both the enactivist and affordance-based approaches to mental disorder that I discuss in this paper acknowledge the role of social norms in shaping our sense of what counts as "disordered." While such accounts are broadly naturalistic in the sense that they conceptualise disorder partly in terms of neurobiological dysfunction, they also acknowledge that concepts of mental health and disorder are bound up with sociocultural norms and values. Indeed, it is important to acknowledge that social values play a role in shaping our conception of disorder, and that stigma, barriers, and discrimination may be more incapacitating than symptoms themselves. Yet we still can hold that such conditions involve

disruptions to neurobiological functioning that make it difficult for subjects to adapt to, and fare well in, their surroundings.

In addition, while the title of this paper refers to "disorders," it is important to acknowledge that the labels we use to refer to these conditions likely don't pick out natural kinds, and that conditions given the same label often manifest in quite heterogeneous ways. My focus is on certain symptoms commonly believed to be indicative of some sort of neurocognitive impairment or disruption to cognitive processing. While such symptoms may appear to have little to do with affectivity, at least at first glance, my central claim is that they centrally involve disruptions to selective attention that can be traced to an affective disturbance. If so, the alleged distinction between disorders of thought and disorders of mood needs to be rethought. Those committed to enactivist or affordance-based accounts of mental disorder, in particular, ought to embrace the claim that there is an important sense in which all mental disorders are affective disorders.

But why have I chosen to focus on enactivist and affordance-based accounts of mental disorder? The first reason for this is simply that such accounts are highly influential, and that more and more theorists have begun to appeal to these conceptual frameworks to account for key symptoms of various mental disorders. The second reason is that both accounts highlight the importance of selective attention and the notion that skilful worldly engagement depends on the ability to highlight some considerations while ignoring others.[1] Given that both accounts also emphasise the importance of affectivity and how it contributes to perception and action, the idea that selective attention depends on affectivity follows quite naturally. This is why I believe that theorists who are committed to these approaches, in particular, should endorse the claim that all mental disorders are affective disorders. Ultimately, however, I suspect that any account of mental disorder that points to disruptions to selective attention and responsiveness will need to acknowledge the important role of affectivity.

2. Enactivism and Disordered Sense-Making

There are a range of different accounts that fall under the label "enactivism." Here, I focus on the version of enactivism sometimes termed "autopoietic," or "autonomic," enactivism, articulated by theorists such as Thompson (2007) and Weber and Varela (2002). These accounts highlight the link between mindedness and biology and emphasise that *mind is in life*. What enactivists call "sense-making" involves embodied action and engagement, and is a matter of being intentionally directed toward the world; thus, it is a mode of gauging meaning that is fundamentally *embodied* and *relational*. At a basic biological level, it can be understood as the process whereby living organisms interpret environmental stimuli in terms of their "vital significance" (Thompson 2007). To regulate and sustain themselves, organisms need to maintain an ongoing exchange of matter and energy with their environment. What counts as a useful resource depends on their structure, needs, and their mode of coupling with their surroundings. Physical and chemical phenomena take on meaning only to the extent that they relate positively or negatively to the "norm of the maintenance of the organism's integrity" (Thompson 2007, 70). Thus, at a basic level, sense-making has an integral *normative dimension* and is directed at survival: normativity arises from the self-production and self-maintenance of a precarious system, and, "through its ongoing individuation, the system intrinsically determines" which interactions support its

1 My claim is not that these approaches are simply interchangeable or that their conceptual vocabularies necessarily amount to the same thing. For a further discussion of the relationship between the two approaches, see Heras-Escribano 2019, Maiese 2021b, and Segundo-Ortin 2020.

continued existence and well-being, or threaten its survival (Buhrmann and Di Paolo 2017, 219). By defining itself, and by distinguishing between self and world, "the organism creates a perspective which changes the world from a neutral place to an *Umwelt* that always means something in relation to the organism" (Weber and Varela 2002, 188). Thus, it is the "restless" character of the metabolic process, and the continuous efforts of the living system to seek interactions with its surroundings to get the necessary energetic and material resources, which pave the way for a discerning perspective or point of view. This is the "existentialist" side of the Thompson's life-mind continuity thesis (Thompson 2011, 41).

Colombetti (2014) has emphasised that the discriminative capacity that allows a living organism to monitor and regulate itself with respect to its conditions of viability has an integral *affective-evaluative* dimension insofar as it involves the living organism being "affected or struck by the suitability of an event for its own purposes" (19). Affectivity appears to be rooted at least in part in the appropriative activity of metabolism and the fact that a living organism's material self-maintenance *matters* to it in some basic sense. Because the primary condition of life is one of continuous self-renewal, an organism's forward trajectory is fuelled by desire: there is a constant need to supply itself with what it lacks to keep going, and "every contact with the world has, for the organisms, an existential meaning" (Weber and Varela 2002, 118). An entity that is concerned with getting the material resources it needs for its own survival and self-maintenance projects this concern onto its surroundings, which is at the core of sense-making. An animal that was incapable of attending selectively to its surroundings would be faced with a potentially endless array of possible stimuli, which would make effective engagement with its environment very difficult. Thus, even at a basic biological level, selective attunement and responsiveness to environmental perturbations centrally involves affectivity. Whether particular environmental stimuli capture an organism's attention has much to do with whether those stimuli have "vital significance."

At a more sophisticated level, sense-making encompasses the various forms of intentional directedness displayed by living animals. Intentionality is generally understood as feature of cognition, perception, and other mental states that are "about" something or "directed at" something. According to enactivism, living animals play an active role in the generation of meaning, and they are intentionally directed towards the world that surrounds them, *in and through* their bodies. In cases of "motor intentionality," for example, embodied agents make sense of objects in relation to themselves, identifying those objects pragmatically, in relation to their goals. Via the formation of bodily habits and associations, they can engage with their surroundings in a bodily and skilful way. Objects and events in their surroundings call for a certain mode of action insofar as agents have a sense that their situation deviates from some optimal body-environment relationship (and that it is contrary to their felt desires). The activity they undertake aims to move them closer to that optimum (thereby bringing the situation more in line with what they desire) (Thompson 2007, 312). Features of the environment that have little bearing on these goals are ignored, while other, more relevant, considerations become the focus of their attention.

Likewise, in perception, an agent interprets incoming stimuli selectively and in relation to pragmatic concerns. They are touched, affected, and stimulated, and what is experienced *matters* in some way or another. What is disclosed is partly a matter of the features of that which exists in the world, and also partly a matter of what they bring to bear, given their specific bodily structure, capacities, and interests. Thus, both selective attention and context sensitivity involve the active participation of the agent in gauging the meanings of target stimuli. Insofar as intentionality is thoroughly bound up with affectivity and what an agent feels is important (Colombetti 2014), it is inherently evaluative. Following, but also extending, Heidegger, one might say that it

is someone's capacity for *affective and practical intentionality*, i.e., "care," which allows them to apprehend the world "as a significant whole, an arena of possible projects, goals, and purposes" (Ratcliffe 2002, 289). Subjects do not passively receive information, but instead play an active role in enacting the meaning, significance, and import of things in their surroundings, in accordance with what they care about. Thus, enactivism emphasises that cognition (sense-making) and affectivity are not clearly distinguishable or separable, but rather *integrated* and *interdependent* (Colombetti 2014). Affectivity focuses attention and guides agents' enactments of meaning.

This enactivist notion of sense-making takes us beyond intentionality and intentional content as traditionally construed; its emphasis on active, embodied engagement allows us to switch focus from intentional *content* to a subject's embodied know-how and their capacity for intentional engagement and responsiveness to their surroundings. This understanding of intentionality, in turn, allows for a new understanding of mental disorders: such conditions do not principally involve disruptions to *knowing-* or *perceiving-that*, but rather difficulties with *gauging-how* to regulate one's coupling with the environment so as to engage effectively with relevant cognitive and practical options. That is, they centrally involve some sort of disruption to *the relation between a living animal and its environment*. Along these lines, de Haan (2017; 2020) maintains that mental disorders should be understood as *disordered patterns of sense-making*. When someone has a disorder, there is a recurring and more or less stable pattern in how someone's sense-making has "gone astray," so that "the way in which the person makes sense of [their] world is biased in a specific direction" (de Haan 2017), and their engagement and responsiveness to the world is distorted. Difficulties with adjusting or attuning their sense-making can result in overly rigid patterns of interaction (de Haan 2020, 10) that undermine their capacity for adaptive engagement, and result in stress, fatigue, dysfunction, or disorder.

This account of mental disorder rests on an important distinction between biological sense-making and what de Haan terms "existential sense-making." Whereas biological sense-making involves making sense of environmental stimuli in terms of their impact on survival and self-maintenance, existential sense-making is underdetermined by biological meanings and values, and instead concerns the way in which people relate to themselves, their world, and other people. And while biological sense-making discloses a valenced environment for living organisms, existential sense-making discloses a much richer, value-imbued world for reflexive beings such as humans (de Haan 2017, 532). This is because existential sense-making involves taking up a stance whereby an agent relates to their self, their experiences, and their situation, and becomes aware of their own patterns of engagement with the world. While this more sophisticated sort of sense-making sometimes involves self-reflection, it also can be more unreflective and implicit in someone's actions. Even if they are not aware of their existential stance or orientation, it influences how they relate to and understand themselves and their experiences. According to de Haan, biological sense-making remains intact in cases of mental disorder; it is not that individuals have trouble satisfying norms of biological self-maintenance or survival. Instead, such conditions centrally involve disruptions to existential sense-making; that is, they are "disorders of the way in which people relate to themselves, their world, and/or other people" (2017, 533).

I have argued (2021a), in contrast, that mental disorders involve disruptions to both biological sense-making and existential sense-making. In my view, disruptions to what de Haan (2017) terms "existential sense-making" are fully bound up with disruptions to living bodily dynamics, including those associated with skin conductance, the respiratory system, and the digestive system. Support for this claim comes from the fact that mental disorders such as depression and schizophrenia centrally involve altered bodily comportment and a loss of bodily attunement. For example, subjects with depression often exhibit diminished bodily activity, sluggishness, lethargy, slouched posture, and tension in the shoulders. Insofar as these disruptions to living

bodily dynamics are fully bound up with disruptions to a subject's ability to make sense of themselves, their world, and other people, disruptions to existential sense-making cannot truly be distinguished from disruptions to biological sense-making.

Despite this disagreement, however, both de Haan and I are committed to the view that mental disorders involve disordered sense-making patterns. Thus, we agree that a mental disorder isn't simply the output or causal result of a neurological dysfunction, but rather a breakdown in the agent's ability to regulate their coupling with the world that surrounds them, including the social world. In a basic biological sense, sense-making has to do with survival and adaptivity; among humans, sense-making also concerns "faring well" in a particular socio-cultural setting. And, as enactivist theorists have emphasised, sense-making is deeply *affective*: "The world takes on significance and value precisely in relation to what the organism is concerned about and striving for" (Colombetti 2014, 19). This seems especially striking in the case of existential sense-making, whereby one makes sense of oneself, one's experiences, and one's situation. The meaning of events and experiences has much to do with what matters to a subject and what they care about; that is, this sort of sense-making necessarily takes place from the standpoint of a *concerned perspective*. Thus, if mental disorders are characterized by disordered patterns of sense-making, it becomes difficult to escape the conclusion that such disorders have an integral affective dimension. In cases of depression, for example, an agent may exhibit heightened attunement to negative considerations, may not attend to relevant positive considerations, and may find it difficult to adjust their sense-making in response to situational factors (Maiese 2021a). These disruptions to sense-making point to a breakdown in selective responsiveness: subjects continue to focus on the negative even when such considerations do not warrant so much attention, and they tend to ignore relevant features of themselves, their surroundings, and other people. Such biased patterns of sense-making are maladaptive insofar as they signify a disruption to selective attunement and responsiveness that makes it difficult for agents to fare well in their surroundings.

3. Affordance Disclosure and Disruptions to Skilled Intentionality

Some theorists have appealed to the theoretical notion of "affordances" to make sense of the disruptions to cognition and affectivity that occur in cases of mental disorder. The notion of affordance is a theoretical concept introduced by J.J. Gibson (1979) that emphasises the complementarity of the animal and the environment, and the link between perception and action. What the environment affords are "what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill" (Gibson 1979, 237). The environment dynamically offers various possibilities for interaction and engagement, but only in relation to an organism with particular capacities. For example, an object is graspable by virtue of its physical properties together with an animal's bodily structure and ability to grasp that object. Thus, affordances can be understood as action possibilities that are specified relationally. However, it appears that animals do not engage with all the action possibilities that are available to them on the basis of their bodily structure and capacities; some affordances that the environment offers are unimportant to a subject because they have no bearing on that subject's goals (Rietveld and Kiverstein 2014).

This has led some theorists to distinguish between available action-possibilities and the smaller subset of affordances that become relevant from the standpoint of a particular agent. The landscape (or "total ensemble") of available affordances is comprised of "the entire set of affordances that are available, in a given environment at a given time" to particular sorts of living organisms (Ramstead, Veissière, and Kirmayer

2016). The *field* of affordances, in contrast, consists of the relevant possibilities for action that a particular individual is responsive to in a concrete situation (Rietveld and Kiverstein 2014). Whereas "the world is rich with affordances . . . only a subset of the possible actions the environment offers matter to a person enough to move them to action" (Kiverstein 2015, 536). Thus, an affordance field can be understood as the "situation-specific, individual "excerpt" of the general landscape of affordances" (de Haan et al. 2013, 7) offered by the environment that stand out as relevant for a particular agent in a specific situation. For an affordance to have relevance is for it to invite the subject to act and beckon certain forms of perceptual-emotional appraisal and bodily engagement (Ramstead, Veissière, and Kirmayer 2016, 4–5).

To make sense of these invitations to act, some theorists have introduced the notion of a "solicitation." Rietveld and Kiverstein (2014), for example, maintain that an affordance becomes a solicitation "when it is relevant to our dynamically changing concerns," takes on a "demand character" and becomes manifest at the bodily level in a state of "action readiness" (342). Along similar lines, Gallagher describes a solicitation as "an affordance that draws an agent to action due to its relevance, or the way that it stands out in the perceptual field" (2018, 722). Whether affordances solicit action is a thoroughly dynamic process, with objects becoming more or less soliciting over the course of a day. This is because "different objects offer possibilities to act, but some of them are more inviting than others" (Dings 2018, 683) at different times. Among the subjective factors that help to shape on agent's field of affordances are that agent's body scheme, capacities, habits, skills (Weichold 2018), and past experience (Rietveld 2012, 213). Which of the many affordances an agent is responsive to in a particular situation also depends significantly on their unique goals, concerns, needs, interests, and preferences (de Haan et al. 2013, 7), including their longer-term intentions, goals, projects, and commitments (Dings 2021).

Thus, it appears that affordances are not simply perceived via the senses, but rather *disclosed*; that is, they are gauged as relevant or selectively attended to, and this process of selective attention relies heavily on affectivity. Those affordances which become relevant are those with affective significance, and the way in which things appear significant to a particular agent is bound up with their occurrent emotions, more diffuse mood, and general background affective orientation. Along these lines, Rietveld (2012) maintains that "the phenomenon of being attracted or drawn by a solicitation can be understood as an emotional perturbation" (213). A specific action-possibility "solicits action (i.e., calls me to act) only when I am responsive to act (i.e., concerned)" (Dings 2018, 687). A book sometimes affords reading, and at other times affords throwing; which of these affordances solicits me to act depends partly on my cares and concerns, and whether I am feeling angry. Some of these action-guiding cares and concerns are diachronic and pertain to ongoing projects, commitments, and long-term goals, whereas others pertain to shorter-term interests.

Depending on someone's overall mood, occurrent emotions, and enduring concerns, some action-possibilities have greater significance or matter more than others, or in different ways. Insofar as an agent's affective attunement to a particular situation helps to determine which available affordances solicit action, this attunement guides their selective responsiveness and enables them to engage effectively with a world that is overflowing with affordances. Along these lines, and building on the work of Merleau-Ponty, Kiverstein and Rietveld (2021) maintain that it "is through the caring engagement of their bodies that subjects experience a meaningful world" (82). It is an agent's concerns, and the way that they emotionally appraise their environment based on these concerns, that render some affordances "inviting and enticing, and others threatening and repelling" (Kiverstein 2015, 536).

It is worth noting that agents typically experience what an object affords or solicits unreflectively, as a result of who they are and what they care about. When a lifelong vegetarian, for example, sees a steak on the barbeque, they require no conscious reflection on their values to ensure they do not see the steak as "affording eating" (Dings 2020b). Instead, this capacity for selective responsiveness frequently operates pre-reflectively, prior to conceptualisation, and involves very fine-grained discriminations connected to what this agent *feels in their gut*, as it were. These affectively contoured discriminations are at work, for example, when someone has a funny feeling about a situation, or spontaneously develops a negative first impression of someone. According to an affordance-based account of selective attention, we need not appeal to mental representations or propositional content to make sense of how this occurs. Instead, it appears that human agents possess some kind of pre-theoretical, non-intellectual understanding of where to direct their attention in a given context, which is built up through learning and mediated by past experience. As they navigate through the world, they do not sequentially process all the information that is potentially available to them, but instead focus on certain very specific things rather than others. Affect operates as the "allure" of consciousness and implies a "dynamic gestalt or figure-ground structure" whereby some objects emerge into affective prominence, while others become unnoticeable (Thompson 2007, 374). The agent's pre-reflective affective attunement to a particular situation, informed by their concerns and interests, thereby helps to determine which available affordances solicit action.

While the prefrontal lobe and other brain regions no doubt play a crucial role, the provision of affective and motivational colour or tone to events and situations is not simply a neural achievement. The affective dimension of selective attention also centrally involves bodily arousal and bodily feelings, which are best understood as distributed over a complex network of neurobiological processes, including metabolic systems, endocrine responses, musculoskeletal changes, and cardiovascular responses. Likewise, selective responsiveness depends significantly on bodily sensitivity, which consists in various changes in heart rate, blood pressure, hormones, skin temperature, and the orientation and positioning of body parts. Associated bodily feelings help to determine the focus of both perception and action by highlighting considerations with *felt importance*. As a result, the very way in which the world is disclosed to the agent is shaped and contoured by bodily feelings of caring. Thus, the process whereby affordances are disclosed as relevant goes beyond mere sensory perception, and involves an affectively laden appraisal process whereby some action possibilities are disclosed as salient and inviting (Dings 2020a).

One might wonder whether individuals can focus their attention on relevant features and considerations simply by relying on conceptual resources (i.e., without the help of affectivity). However, the Jastrow duck-rabbit phenomenon shows that the mapping from attention to the natural stimulus can even be uniformly underdetermined across our species, including all cognisers who possess the concepts DUCK, RABBIT, and PICTURE. It is because agents attend to different aspects of a visual stimulus that they may see either the duck or the rabbit. What is more, some degree of selective attention takes place in all our interactions with our surroundings, though we tend not to notice it. Affordance-based accounts emphasise that selective attention often centres around fine-grained, action-oriented adjustments to behaviour, and that feelings of subjective import are central to this process. Note that this is fully consistent with phenomenological observations. In our everyday lives, the action possibilities that we experience as relevant, and which grab our attention, are those that reflect our concerns, emotions, moods, background affective orientations, and enduring commitments.

Many theorists have looked to these notions of affordance, solicitation, and affordance disclosure to make sense of the disruptions to functioning that occur in mental disorder. For example, de Haan et al. (2013) point to a

breakdown in the structure of the field of relevant affordance, which involves at least three key dimensions: (1) the "width" of the field refers to the broadness of the scope of affordances that an individual perceives, which relates to their having a choice or options for action; (2) the "depth" of the field refers to the temporal aspect: the individual not only perceives affordances that are immediately present here and now, but is also pre-reflectively aware of future plans and action-possibilities; and (3) the "height" of the affordances in the field refers to the relevance or importance of the affordances that one is responsive to, i.e., "the experienced solicitation or affective allure" (de Haan et al. 2013, 7), and relates to salience and motivation. In cases of mental disorder, the width, depth, or height of an agent's affordance field become narrower, lack depth, or flatten out. And because engagement with relevant affordances is a function of the subject's cares and concerns and their sense of who they are, alterations to the affordance space typically also involve modifications to self-experience.

Similarly, Gallagher (2018) characterises different disorders in terms of distinctive changes in the affordance space. He says that this space is defined by evolution and bodily structure (e.g., the fact that an agent has hands), development (e.g., the agent's life-stage, and whether they are an infant or an adult), social and cultural practices (including normative constraints), and the subject's past experience, skill level, and education. Thus, if I lack the sensorimotor skills and capacities needed to climb the cliff in front of me, cliff-climbing is not part of my affordance space. In the case of depression, there are changes in a subject's affordance space: because drive, impulse, and appetite are reduced or lost, many action-possibilities are closed off. Likewise, in cases of schizophrenic delusion, affordances change: a chair might appear not as something to sit in, but rather as a "thing" that has lost its name, function, and meaning. And, in the case of OCD, "there is a serious constriction of the affordance space as the subject finds herself limited to one repetitive action, or one set of specific actions, and unable to move beyond that" (723). Presented with a hammer, the subject is unable to refrain from picking it up and using it. According to Gallagher, "particular types of psychiatric disorder reorient or reorganize concerns, interests, and abilities, and thereby change what counts as an agent's affordances or solicitations" (725). Similarly, in their description of changes in someone's physical-mental-affective health, Ramstead, Veissière, and Kirmayer (2016) suggest that adjustments across the affordance space can re-sculpt "a field of solicitations out of the total landscape of available affordances," dynamically moving the organism toward transformations in what counts as an optimal grip in a particular situation (13). While these different accounts differ in terms of their details, they all emphasise that disruptions to affordance engagement stem from an inability to gauge relevance or to gain an optimal grip on available action-possibilities.

This notion of "optimal grip" also plays a central role in the Skilled Intentionality Framework (Rietveld 2012; Rietveld et al. 2018; van Dijk and Rietveld 2017). Rietveld (2012) characterises the normative aspect of engagement with affordances in terms of embodied know-how and skill: "Acting appropriately requires that a complex and particular situational context is taken into account by the individual's motor intentional activity" (215). Taking into account their particular situation, capabilities, and interests, the agent regulates themselves and coordinates their engagement with the environment so as to move toward "optimal grip on multiple relevant affordances simultaneously, that is, on a field of relevant affordances" (Rietveld et al. 2018, 45). This is a matter of being appropriately responsive to available affordances, being able to switch from one sort of activity to another as events unfold over time, being open to engagement with previously unexplored affordances, and being able to modify built-up patterns of engagement. What Dreyfus (2007), building on the work of Merleau-Ponty, describes as "skillful coping" requires that "embodied beings like us take as input energy from the physical universe and respond in such a way as to open them to a world organized in terms of their needs, interests, and bodily capacities" (251). Adaptive agency, then, "can be understood as an agent's having a grip on a rich, dynamic, and varied field of relevant affordances" (Ramírez-Vizcaya and Froese 2019, 8); it involves

"a constant on-going sensorimotor loop of perceiving invitations for action, acting on them, perceiving new invitations, and so on" (Weichold 2018, 772). Even so-called "higher" forms of cognition, such as imagination, long-term planning, perspective-taking, and decision-making can be understood in terms of skilled activities of engaging with worldly situations and coordinating with multiple relevant affordances simultaneously.

In cases of mental disorder, agents are unable to achieve this "optimal grip" on available affordances. Whereas bodily affectivity ordinarily helps agents to engage selectively with specific features of their surroundings and thereby become attuned to relevant action-possibilities, mental disorder disrupts this capacity for selective attention. However, the difficulty is not simply that someone's sense organs are malfunctioning, rendering them unable to "perceive" objects as issuing demands. Rather, the agent faces challenges with respect to gauging the relevance of available affordances and being solicited appropriately. Disruptions to affordance disclosure take a variety of forms, depending on the disorder (Maiese 2021b). In some cases, the central difficulty is that subjects are solicited by irrelevant affordances. In other cases, relevant affordances go unnoticed. And, in others, the affordances that solicit action lack a sense of "mineness." In my view, all these disruptions can be understood as some sort of disruption to selective attention and responsiveness, rooted in an affective disturbance. But what about the language disturbances sometimes found in schizophrenia and the "context blindness" sometimes exhibited by autistic subjects?

4. Language Disturbances and "Context Blindness"

Perhaps few people will object to the claim that key symptoms of major depression or bipolar disorder result from disruptions to affectively grounded selective attention processes. However, it may seem less obvious that symptoms involving language and contextual integration can be accounted for in these terms. Whereas these symptoms are commonly understood to be the result of some sort of neurocognitive deficit, I argue that they are better understood in terms of disordered sense-making and a diminished ability to engage effectively with available affordances.

Language and affectivity often have been viewed as belonging to two distinct domains that must be studied separately. Whereas language belongs to structures of thought, is based on words and representations, and is communicatively deliberate, affectivity belongs to the body and is associated with unintentional sensations and reactions (Jensen 2014, 1). These sorts of assumptions may very well lead some theorists to balk at the suggestion that language difficulties are the result of an affective disturbance. But do disruptions to affectively guided selective attention processes contribute to such impairments? Such difficulties include repetition of phrases, frequent uncompleted sentences, neologisms, and "word salad."[2] Consider the following example from Saks (2007): "I'm just kidding around. . . . Kidding has to do with sheep. I'm sheepish. Have you ever killed anyone? I've killed lots of people with my thoughts" (215). In this example, single elements of language lose their function as carriers of intentional meaning (Fuchs and Röhricht 2017, 132) and stand out separately from the overall gist of the sentence. As a result, subjects are inadequately responsive to relevant linguistic affordances.

To explain why word intrusions disrupt the speech of some subjects with schizophrenia, Maher (2003) points to "defective employment of inhibitory activity necessary to exclude intrusions" (19). The ability to speak a sequence of words in a sentence is made possible by the ability to inhibit irrelevant associations for each

2 For a review, see Covington et al. 2005 and Docherty, DeRosa, and Andreasen 1996.

separate word, as well as the ability to screen out external sources such as background conversation. That is, to focus their attention on a limited and significant aspect of language input, the subject must actively inhibit word associations that are not relevant given their concerns or the situation at hand. However, because subjects with schizophrenia are deficient in these inhibition mechanisms, and find it difficult to focus their attention on relevant contextual features, they are highly susceptible to word associations that are not relevant to the case at hand. For example, there is an association between the word "orange" and the word "Florida," but this association will be relevant only in particular contexts, in which specific interests, cares, or concerns are at play. When irrelevant word associations (ones that do not reflect what the subject cares about) intrude and solicit speech, they disrupt a subject's ability to complete meaningful sentences. A subject's overt utterance may not be in line with the thought that prompted it or the general background of their cares and concerns, making it difficult for them to communicate effectively with their interlocutors.

The notion that disruptions to affectivity and selective attention help to account for some of these key symptoms seems to resonate deeply with both enactivist and affordance-based approaches. First, some enactivist theorists have emphasized that language use closely reflects and expresses a subject's concerns. Cuffari (2014), for example, suggests that someone's particular use of language involves a particular way of caring and being careful in their speaking to others; that is, their linguistic sensitivities and habituated patterns of expressivity and response help to comprise their sources of caring and evaluating (Cuffari 2014). If so, then disruptions to affectivity would contribute to a kind of inability to be properly careful or attentive in one's employment of language. And in their discussion of skilful engagement with affordances, Kiverstein and Rietveld (2021) maintain that "the bodily activity of speaking is expressive of care – the ways of being involved with the world manifest in the regular patterns of doing in the speaker's form of life" (187). I hypothesise that affective disturbances make it difficult for some subjects with schizophrenia to inhibit irrelevant linguistic input, screen out unneeded or irrelevant words or phrases, and engage effectively with "linguistic affordances." In such cases, both relevant and irrelevant affordances have the same "height," and none of them stand out as especially salient. As a result, the subject is solicited to speak by language affordances that are not relevant given their cares and interests, and they find it difficult to determine which of the several meanings of a word is relevant for them in the situation at hand. What is more, due to their inability to gauge relevant speech affordances, a subject with schizophrenia may retreat into their own world; this further detracts from their ability to engage effectively with available action-possibilities.

It is also worth highlighting that language disturbances are not the only symptoms of schizophrenia that are indicative of a disruption to selective attention. Subjects also commonly exhibit deficits in perceptual grouping, so that objects do not stand together in an overall context and instead appear as meaningless details. They also have difficulty excluding distracting visual, auditory, and tactile input when trying to concentrate on selected parts of the environment (Maher 2003, 14). All these tasks require the inhibition of irrelevant information and the ability to focus one's attention on relevant considerations. However, among subjects with schizophrenia, there is a notable discrepancy between the amount of attention something deserves and the amount that it receives. What is more, in cases of so-called "unworlding," subjects experience a sense of strangeness about external objects that ordinarily would seem familiar, and subjects find themselves less capable of engaging with and "grasping" their surroundings (Sass 2004). Insignificant details of the surrounding environment become conspicuously salient, commonplace objects seem to lose their familiar meaning and recognizable significance, and the cognitive or perceptual world undergoes a kind of fragmentation. A chair, for example, might appear not as something that affords sitting, but rather as a "thing" that has lost its function and meaning.

Such considerations indicate that subjects with schizophrenia exhibit disordered sense-making and encounter difficulties with affordance engagement across a wide range of domains. This includes selective attention deficits that impact their ability to process linguistic stimuli, remember events, engage interpersonally with others, and exhibit executive control. These difficulties with selective attention are commonly viewed as sources of cognitive dysfunction in schizophrenia (Gold et al. 2018). However, the claim that these difficulties with selective attention result, in part, from an affective disturbance is supported by the fact that subjects with schizophrenia commonly experience diminished bodily attunement and strange bodily affective states. The quasi-affective sensations and bodily states that subjects experience include "sensations of movement or pulling or pressure inside the body or on its surfaces; electric or migrating sensations; awareness of kinaesthetic, vestibular, or thermic sensations; and sensations of diminution or enlargement of the body or its parts" (Sass 2004, 135). Thus, disruptions to selective attention occur alongside disruptions to overall affective bodily attunement. But why think there is a link between the two?

I propose that, because their framework of bodily attunement is diminished, subjects begin to experience sensations that are dissociated from their ongoing sense of self, lack a sense of personal relevance, and are experienced as free-floating rather than being meaningfully directed toward the world. In fact, there is evidence that subjects with schizophrenia commonly exhibit blunted affect and deficits in emotional expressiveness (Trémeau 2006). Kring and Moran's (2008) review of emotional response deficits in schizophrenia found that subjects are less expressive (both facially and vocally) than individuals without schizophrenia in response to a variety of evocative stimuli. For example, individuals with schizophrenia appear to display fewer positive and negative facial expressions in response to emotionally evocative film clips, foods, and social interactions (Kring and Moran 2008, 821). Yet, even though these individuals are markedly less expressiveness, they do not seem to differ as much with respect to their reports of emotional experience or their affective arousal (as measured by skin conductance reactivity, for example). Such data indicate that, although subjects with schizophrenia do experience emotion, it does not guide their behaviour in the usual way, leading to disruptions to agency and a decreased ability to gauge the relevance of available motor and social affordances.

Like subjects with schizophrenia, autistic subjects appear to encounter difficulties with selective attention. For example, they often say things that lack relevance to the situation at hand, are less able to deceive others, and often do not produce most aspects of pragmatics in their speech. They appear to have a limited understanding of the relationship between language and social context, and find it difficult to interpret a speaker's words in relation to their perspective and intentions. Because they do not make appropriate use of context to interpret the communicative intentions of their interlocutors (Vermeulen 2015, 186), they have difficulty understanding metaphor, sarcasm, and irony, and tend to adopt a literal interpretation of these figurative modes of speech. They are less capable of homing in on social and contextually-based meanings that go beyond the semantic meaning of what is said. For example, autistic subjects have difficulty with using context to disambiguate the pronunciation or meaning of homographs (Lopez and Leekam, 2003). These are cases in which, in order to choose the correct (i.e., contextually appropriate) pronunciation of a word, one must understand that word in relation to the whole sentence's meaning. Because contextual disambiguation is difficult for autistic subjects, they tend to produce the word's more common pronunciation, regardless of the preceding sentence's meaning and overall context (Frith and Happé 1994, 124). They also exhibit deficits with respect to the completion of sentences with words that are congruent with the entire sentence (Booth and Happé 2010). For them, reading a sentence is more like reading a list of unconnected words. What is more, they tend to violate informal dialogical rules such as turn-taking, and often continue talking about a single topic regardless of whether it is relevant to the hearer. Such considerations lead Stanghellini (2001) to suggest that the autistic person's use of

language is much like a soliloquy rather than a "cooperative process whose aim is interlacing one's own world with that of others" (296). This points to a breakdown in the ability to engage with relevant linguistic and social affordances.

Autistic subjects also have difficulty integrating perceptual information, often show a preoccupation with details and parts, and fail to extract gist or configuration. Their difficulties with selectively attending to salient contextual features surfaces in their insusceptibility to certain perceptual illusions, their knack for finding embedded figures within a larger design, and their difficulties with "gestalt" perception (seeing whole figures or scenes as opposed to their parts). For example, children with autism excel at the Embedded Figures Test, which involves spotting a hidden figure among a larger design or meaningful drawing. In addition, they tend to have a knack for locating tiny objects (such as thread on a patterned carpet) and are good at detecting minute changes in familiar layouts, such as the arrangement of items on a bathroom shelf (Frith and Happé 1994, 122). And, when asked to make judgements about standard textbook visual illusions, they tend not to succumb to illusions as much as non-autistic subjects do. Because autistic subjects focus on the to-be-judged parts without integrating them with the surrounding illusion-inducing context, they are better able to make accurate judgements in such cases. However, subjects who have difficulty gauging the contextually appropriate meanings of perceptual scenes will be less capable of making sense of their surroundings and evaluating, disambiguating, and changing between several different possible interpretations in light of contextual considerations (Vermeulen 2015, 185). This is likely to result in rigid and absolute links between stimuli and meanings, and thus a lack of behavioural flexibility and insufficient attunement to relevant affordances. Decreased contextual sensitivity likely plays a role in the difficulties autistic subjects sometimes have in distinguishing the important from the incidental and attending to the salient stimuli in a given situation (Loth, Gomez, and Happé 2011). Due to a lack of selective bias towards relevant aspects of perceptual input, autistic subjects often treat all perceptual details as equally important; as a result, cognitive processes can become "bogged down in minute details or irrelevancies" (Vermeulen 2015, 184). Subjects may pay attention to things that should be ignored, given the context, their immediate interests, and their longer-term goals.

To make sense of this sort of "context blindness," some theorists have suggested that autistic subjects exhibit "weak central coherence" (Frith 1989). The term "central coherence" refers to "the natural human tendency to draw together several pieces of information to construct higher-order meaning in context" (Frith and Happé 1994, 121). The notion that autistic subjects exhibit "context blindness" builds upon the notion of weak central coherence, and stresses that context use plays a central role in gauging relevance and guiding attention (Vermeulen 2015, 182). The notion of context blindness in autistic spectrum disorder refers primarily to difficulties with using context when making sense of one's surroundings and other people, i.e., in using context to interpret what is perceived (Vermeulen 2015, 183). Along these lines, Morsanyi, Handley, and Evans (2010) described the autistic mind as "decontextualized." After correctly naming the bed, mattress, and quilt, an autistic boy referred to the pillow on the bed as "a piece of ravioli" (Happé 1994, 118). The fact that he correctly named these other objects suggests that he "saw" the context of the pillow, but that contextual considerations did not appropriately feed into his interpretation of the meaning of this object. This reduced contextual sensitivity may contribute to the rigid or straightforward thinking often seen among autistic subjects. Indeed, since autistic subjects often rely on fixed meanings and rules, and find it difficult to alter their behaviour in response to contextual considerations, their sense-making is biased in a particular direction (de Haan 2020) and they find it difficult to obtain optimal grip on available affordances.

Yet, despite this striking evidence of decreased contextual sensitivity, Vermeulen (2015) notes that "we still lack a clear understanding of the difficulties that people with [autism spectrum disorder] have in perceiving and using context" (Vermeulen 2015, 189). To make sense of this, I have proposed that the ability to utilise context in various forms of sense-making and perception depends significantly on affectivity and bodily-affective attunement. Which contextual considerations are relevant depends significantly on an agent's immediate concerns and longer-term goals, including their concerns about satisfying existing social norms and expectations. Consider that most of the phenomena that we perceive and engage with in the environment can be construed in different ways and are open to multiple interpretations. Affectivity helps us to derive meaning from a world of objects, events, and other people whose appearance, workings, or behaviours are inherently ambiguous. Likewise, gauging global meaning requires interpreting details against a background of significance. Without the direction and focus of attention provided by bodily affectivity, autistic subjects find it difficult to ascertain the broader meaning of things and may focus instead on superficial or irrelevant details. As a result, they find it easier to screen out context (in the case of finding embedded figures within a larger design), but also struggle to see "the big picture" as opposed to its parts. Whereas subjects ordinarily interpret perceptual stimuli from the environment on the basis of their cares and concerns, the sense-making of autistic subjects is less affectively contoured. Without a stable framework of bodily affectivity to serve as a guide for sense-making and affordance perception, autistic subjects encounter difficulties navigating the material and social world.

What is more, subjects with autism often exhibit what Krueger (2021) terms "style blindness": they have a decreased ability to gauge the style (or the "how") of various forms of movement. Here Krueger appeals to Daniel Stern's (2010) notion of "forms of vitality," which encompass the manner or style in which an action is executed. For example, a particular movement sequence (e.g., chopping up vegetables for dinner) might be done in a confident way, an angry way, or a lethargic way. Whereas children with autism tend not to struggle with imitating the "what" of an action (nor carrying out complex, goal-directed movement that matches the movement of the person they are imitating), they do display deficits with respect to imitating the style (the "how") of this movement. Such evidence suggests that subjects with autism find it difficult to gauge the expressive qualities of various forms of movement; as a result, they may find it difficult to make sense of others' behaviour, gauge relevant social affordances, or respond appropriately to others' actions. In my view, this sort of "style blindness" is directly connected to their diminished affective bodily attunement, which renders them less capable of detecting the highly fine-grained expressive features of others' actions.

The claim that these difficulties stem, in part, from an affective disturbance is supported by evidence showing that the bodily arousal associated with affective states makes an essential contribution to selective attention. Pessoa (2008) explores how cognitive and affective processing are integrated in the brain and claims that the cognitive and emotional contributions to executive control conjointly and equally contribute to the control of thought and behaviour. Likewise, Panksepp (1998) describes emotion as a collection of meaning-generating and adaptive mechanisms that are rooted in specific neural and endocrine processes, and which allow the organism to adapt to life-challenging circumstances. And Lewis (2005) discusses how the sub-personal processes that underlie appraisal and emotion are a distributed network of self-organising and mutually influencing brain and bodily processes. Together with the amygdala, bodily arousal and endocrine activity help to maintain an organism's homeostatic equilibrium, enhance attention, and prepare the individual for action. Likewise, the work of Barrett and Bar (2009) suggests that bodily feelings and sensations signalling an object's salience or relevance assist in perception and object recognition from the very moment that visual stimulation begins. Even among very young humans, there is a link between emotional response,

movement, and the capacity to distinguish between different kinds of sensory stimuli (Ciompi 2003). Indeed, it is widely accepted among neuroscientists that "rapid and efficient selection of emotionally salient or goal-relevant environmental stimuli is crucial for flexible and adaptive behavior" (Yamaguchi and Onoda 2012, 1). But, whereas some theorists might account for context insensitivity in reductive terms and point to brain dysfunction, the enactivist and affordance-based approaches I have discussed in this paper would predict disruptions to the whole living body and the way in which an agent as a whole relates to and actively engages with their world.

In fact, existing research supports the notion that autistic subjects exhibit disruptions to overall bodily attunement. A study by Leekam et al. (2007) found that over 90% of autistic children exhibit sensory abnormalities. First-person autobiographical accounts often reveal that subjects have unusual sensory experiences, including insensitivity to pain and atypical responses to auditory, visual, tactile, and olfactory stimuli. For example, subjects may become distressed or unusually fascinated by certain sounds, exhibit an unusual degree of interest in bright lights or shiny objects, or exhibit a negative reaction to gentle touch. Research also indicates that autistic children have difficulty replicating facial expressions and fail to exhibit gestural emotion in peer interactions (Langdell 1981). Marcari et al. (2018) found that toddlers with ASD exhibited a muted affective response to novel, intrusive stimuli, and Yirmiya et al. (1989) found that Children with ASD displayed significantly less affect in their facial expressions compared to a control group of children with developmental delay. Macdonald et al. (1989) demonstrated that emotional expression deficits extend into adulthood, even for high functioning adults, and that subjects were relatively impaired with respect not just to the appreciation of the emotions of others, but also the production of emotional expressions. This points to a disruption to affectivity, which I have suggested contributes to the "context blindness" commonly found among autistic subjects.

5. Conclusion

I have argued that difficulties with selective attention and contextual sensitivity are commonly found in both schizophrenia and autism, and that these difficulties stem partly from disruptions to affective bodily attunement. How, then, can we differentiate between the symptoms found in these conditions? Vermeulen (2015) speculates that perhaps not being able to activate contextually appropriate meanings is especially typical in autism, whereas not being able to repress contextually inappropriate meanings is more characteristic of schizophrenia (188). Another possibility is that these deficits are quite similar in terms of their underlying causes and dynamics, and that what principally differentiates schizophrenia and autism is the presence or absence of yet other symptoms. There remains work to be done to provide an enactivist or affordance-based account of these disorders, and to examine how disruptions to affectivity and selective attention contribute to key symptoms.

A lingering question concerns the intended scope of the argument presented here. Have I adequately defended the claim that all mental disorders are affective disorders? My central (relatively modest) aim has been to show that proponents of enactivist and affordance-based accounts should embrace this claim. But what about those who are not committed to such accounts? As noted earlier, if it is true that the capacity for selective attention depends significantly on affectivity, then any account of mental disorder that highlights the role of selective attention will need to acknowledge this. Of course, a full defence of the claim that all mental disorders are affective disorders might require that I discuss a much wider range of mental disorders and characteristic symptoms; there is not enough space to undertake such a defence here.

But what about delusions, such as thought insertion, which are sometimes found in schizophrenia? Since such delusions are commonly thought to be the result of impaired rationality or diminished epistemic capacities, why think that they result partly from some sort of affective disturbance? Consider that, in addition to experiencing possibilities to walk and kick, agents experience possibilities to attend, imagine, and deliberate (McClelland 2019, 170). Which of these *mental affordances* are appraised as relevant (and which possible acts of thinking are solicited) depends partly upon an agent's interests, concerns, and overall affective attunement. This bodily attunement allows for the sort of "automatic proto-assessment" (McClelland 2019, 166) whereby agents select from a shortlist of the options for mental action that they feel, in-and-through their bodies, are more relevant or inviting. Only a fraction of possible thoughts will be solicited. Because trains of thoughts are ordinarily focused and guided by built-up patterns of affective bodily attunement, they do not seem to appear out of nowhere. Among some subjects with schizophrenia, however, the patterns of attention and bodily orientation that ordinarily serve as the backdrop for thought are attenuated. Because solicited thoughts do not arise against the structure-giving backdrop of an agent's desiderative feelings, their concerns about the future, or their current needs and desires, they seem alien and out of context. Without some framework in which particular mental affordances can take on relevance and significance, the thoughts that occur may seem object-like, and the experience of thinking may lose its desiderative tone and world-directedness. Some subjects then attempt to *recontextualize* them by attributing them to some other source (Martin and Pacherie 2013), thereby giving rise to reports of thoughts that are *not theirs*, but nonetheless *in their minds*. While this account is only a sketch, it begins to shed light on how symptoms such as delusions might be understood as resulting from a disruption to affectively driven processes of affordance engagement.

What implications does my proposed account have for treatment? Currently, medication is the most common mode of treatment for many mental disorders. Although I do not advocate the elimination of drug-based therapy, or deny that it ever can prove effective, it is evident that drugs often have negative side effects (in the case of schizophrenia), and also that we do not know of drugs that help to improve cognitive and affective functioning (in the case of autism). What is more, it is unlikely that the mere provision of social skills training will be enough to teach subjects to focus on relevant contextual features, switch course in response to changing circumstances, or navigate smoothly through complex social environments (Vermeulen 2015, 188). Instead, alongside efforts to create more hospitable environments (Krueger and Maiese 2018), we will need to cultivate subjects' overall bodily-affective attunement and capacities for selective attention. Indeed, the argument presented here supports the further development and utilisation of methods that target subjects' emotions and bodily-affective feelings in order to foster adaptive sense-making and affordance engagement. One way to do this is through the development of self-narratives that weave together different elements of a person's life experiences and diachronic concerns (Dings 2018). In addition, further research should be done to investigate the efficacy of expressive arts therapies that involve dance, movement, music, and visual art (Maiese 2015, chap. 6). The potential effectiveness of expressive arts therapies derives from the fact that they engage emotions and bodily-affective feelings to bring about shifts in higher-level cognitive and interpersonal functioning. By tapping in to subjects' bodily affective orientations and attuning them to how their bodies are feeling, expressive arts interventions can strengthen subjects' capacities for selective attention and responsiveness, helping them to make sense of their surroundings and to gain an "optimal grip" on relevant affordances.

References

Barrett, L. F. and M. Bar. 2009. "See it with Feeling: Affective Predictions during Object Perception." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1521): 1325–34.

Booth, R. and F. Happé. 2010. "'Hunting with a Knife and … Fork': Examining Central Coherence in Autism, Attention Deficit/Hyperactivity Disorder, and Typical Development with a Linguistic Task." *Journal of Experimental Child Psychology* 107 (4): 377–93.

Bortolan, A. and A. Salice. 2018. "Introduction: Special Issue: Philosophical Perspectives on Affective Experience and Psychopathology." *Discipline Filosofiche* (2): 5–20.

Buhrmann, T. and E. Di Paolo. 2017. "The Sense of Agency – A Phenomenological Consequence of Enacting Sensorimotor Schemes." *Phenomenology and the Cognitive Sciences* 16 (2): 207–36.

Chapman, R. 2020. "Defining Neurodiversity for Research and Practice." In *Neurodiversity Studies: A New Critical Paradigm*, edited by H. Rosqvist, N. Chown, and A. Stenning, 218–20. Routledge.

Ciompi, L. 2003. "Reflections on the Role of Emotions in Consciousness and Subjectivity, from the Perspective of Affect-Logic." *Consciousness & Emotion* 4 (2): 181–96.

Colombetti, G. 2014. *The Feeling Body: Affective Science Meets the Enactive Mind*. MIT Press.

Covington, M. A., C. He, C. Brown, L. Naçi, J. T. McClain, B. S. Fjordbak, J. Semple, and J. Brown. 2005. "Schizophrenia and the Structure of Language: The Linguist's View." *Schizophrenia Research* 77 (1): 85–98.

Cuffari, E. C. 2014. "Keep Meaning in Conversational Coordination". *Frontiers in Psychology* 5: 1–5.

de Haan, S. 2017. "The Existential Dimension in Psychiatry: An Enactive Framework." *Mental Health, Religion & Culture* 20 (6): 528–35.

———. 2020. *Enactive Psychiatry*. Cambridge University Press.

de Haan, S, E. Rietveld, M. Stokhof, and D. Denys. 2013. "The Phenomenology of Deep Brain Stimulation-Induced Changes in OCD: An Enactive Affordance-Based Model." *Frontiers in Human Neuroscience* 7 (653): 1–14.

Dings, R. 2018. "Understanding Phenomenological Differences in how Affordances Solicit Action. An Exploration". *Phenomenology and the Cognitive Sciences* 17 (4): 681–99.

———. 2020a. "Psychopathology, Phenomenology and Affordances." *Phenomenology and Mind* 18: 56–66.

———. 2020b. "Can Ecological Psychology Account for Human Agency and Meaningful Experience?" *Constructivist Foundations* 15 (3): 220–22.

———. 2021. "Meaningful Affordances." *Synthese* 199: 1855–75.

Docherty, N. M., M. DeRosa, and N. C. Andreasen. 1996. "Communication Disturbances in Schizophrenia and Mania." *Archives of General Psychiatry* 53 (4): 358–64.

Dreyfus, H. 2007. "Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian." *Philosophical Psychology* 20 (2): 247–68.

Frith, U. 1989. *Autism: Explaining the Enigma*. Basil Blackwell.

Frith, U. and F. Happé. 1994. "Autism: Beyond 'Theory of Mind'." *Cognition* 50 (1–3): 115–32.

Fuchs, T. and F. Röhricht. 2017. "Schizophrenia and Intersubjectivity: An Embodied and Enactive Approach to Psychopathology and Psychotherapy." *Philosophy, Psychiatry, & Psychology* 24 (2): 127–42.

Gallagher, S. 2018. "The Therapeutic Reconstruction of Affordances." *Res Philosophica* 95 (4): 719–36.

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin.

Gold, J. M., B. Robinson, C. J. Leonard, B. Hahn, S. Chen, R. P. McMahon, and S. J. Luck. 2018. "Selective Attention, Working Memory, and Executive Function as Potential Independent Sources of Cognitive Dysfunction in Schizophrenia." *Schizophrenia Bulletin* 44 (6): 1227–34.

Graham, G. 2013. *The Disordered Mind: An Introduction to Philosophy of Mind and Mental Illness*. Routledge.

Happé, F. 1994. *Autism: An Introduction to Psychological Theory*. UCL Press.

Heras-Escribano, M. 2019. "Pragmatism, Enactivism, and Ecological Psychology: Towards a Unified Approach to Post-Cognitivism." *Synthese* 198 (Suppl 1): 337–63.

Jensen, T. W. 2014. "Emotion in Languaging: Languaging as Affective, Adaptive, and Flexible Behavior in social Interaction." *Frontiers in Psychology* 5: 720.

Kiverstein, J. 2015. "Empathy and the Responsiveness to Social Affordances". *Consciousness and Cognition* 36: 532-42.

Kiverstein, J. and E. Rietveld. 2021. "Scaling-up Skilled Intentionality to Linguistic Thought". *Synthese* 198 (1): 175–94.

Kring, A. M. and E. K. Moran. 2008. "Emotional Response Deficits in Schizophrenia: Insights from Affective Science." *Schizophrenia Bulletin* 34 (5): 819–34.

Krueger, J. 2021. "Enactivism, Other Minds, and Mental Disorders." *Synthese* 198 (Suppl 1): 365–89.

Krueger, J. and M. Maiese. 2018. "Mental Institutions, Habits of Mind, and an Extended Approach to Autism." *Thaumàzein: Rivista di Filosofia* 6: 10–41.

Langdell, T. 1981. "Face Perception: An Approach to the Study of Autism." PhD thesis, University of London.

Leekam S. R., C. Nieto, S. J. Libby, L. Wing, and J. Gould. 2007. "Describing the Sensory Abnormalities of Children and Adults with Autism". *Journal of Autism and Developmental Disorders* 37 (5): 894–910.

Lewis, M. 2005. "Bridging Emotion Theory and Neurobiology through Dynamic Systems Modeling." *Behavioral and Brain Sciences* 28 (2): 169–94.

Lopez, B. and S. R. Leekam. 2003. "Do Children with Autism Fail to Process Information in Context?" *Journal of Child Psychology and Psychiatry* 44 (2): 285–300.

Loth, E., J. C. Gómez, and F. Happé. 2011. "Do High-Functioning People with Autism Spectrum Disorder Spontaneously Use Event Knowledge to Selectively Attend to and Remember Context-Relevant Aspects in Scenes?" *Journal of Autism and Developmental Disorders* 41 (7): 945–61.

Macari, S., L. DiNicola., F. Kane-Grade, E. Prince, A. Vernetti, K. Powell, S. Fontenelle IV, and K. Chawarska. 2018. "Emotional Expressivity in Toddlers with Autism Spectrum Disorder". *Journal of the American Academy of Child & Adolescent Psychiatry* 57 (11): 828–36.

Macdonald, H., M. Rutter, P. Howlin, P. Rios, A. Le Conteur, C. Evered, and S. Folstein. 1989. "Recognition and Expression of Emotional Cues by Autistic and Normal Adults." *Journal of Child Psychology and Psychiatry* 30 (6): 865–77.

Maher, B. 2003. "Schizophrenia, Aberrant Utterance and Delusions of Control: The Disconnection of Speech and Thought, and the Connection of Experience and Belief." *Mind & Language* 18 (1): 1–22.

Maiese, M. 2015. *Embodied Selves and Divided Minds*. Oxford University Press.

———. 2021a. "An Enactivist Reconceptualization of the Medical Model." *Philosophical Psychology* 34 (7): 962–88.

———. 2021b. "Enactivism, the Field of Affordances, and Mental Disorder." *The Journal of Mind and Behavior* 42 (2): 175–96.

Martin, J.-R. and E. Pacherie. 2013. "Out of Nowhere: Thought Insertion, Ownership and Context-Integration." *Consciousness and Cognition* 22 (1): 111–22.

McClelland, T. 2019. "Representing our Options: The Perception of Affordances for Bodily and Mental Action." *Journal of Consciousness Studies* 26 (3–4): 155–80.

Morsanyi, K., S. J. Handley, and J. Evans. 2010. "Decontextualised Minds: Adolescents with Autism are Less Susceptible to the Conjunction Fallacy than Typically Developing Adolescents." *Journal of Autism and Developmental Disorders* 40 (11): 1378–88.

Panksepp, J. 1998. "The Periconscious Substrates of Consciousness: Affective States and the Evolutionary Origins of the Self." *Journal of Consciousness Studies* 5 (5–6): 566–82.

Pessoa, L. 2008. "On the Relationship between Emotion and Cognition." *Nature Reviews Neuroscience* 9 (2): 148–58.

Ramírez-Vizcaya, S. and T. Froese. 2019. "The Enactive Approach to Habits: New Concepts for the Cognitive Science of Bad Habits and Addiction." *Frontiers in Psychology* 10: 301.

Ramstead, M. J. D., S. P. L. Veissière, and L. J. Kirmayer. 2016. "Cultural Affordances: Scaffolding Local Worlds through Shared Intentionality and Regimes of Attention." *Frontiers in Psychology* 7: 1090.

Ratcliffe, M. 2002. "Heidegger's Attunement and the Neuropsychology of Emotion." *Phenomenology and the Cognitive Sciences* 1 (3): 287–312.

———. 2005. "The Feeling of Being." *Journal of Consciousness Studies* 12 (8–9): 43–60.

Rietveld, E. 2012. "Bodily Intentionality and Social Affordances in Context." In *Consciousness in Interaction: The Role of the Natural and Social Context in Shaping Consciousness*, edited by F. Paglieri, 207–26. John Benjamins Publishing.

Rietveld, E., D. Denys, and M. van Westen. 2018. "Ecological-enactive Cognition as Engaging with a Field of Relevant Affordances: The Skilled Intentionality Framework (SIF)". In *The Oxford Handbook of 4E Cognition*, edited by A. Newen, L. de Bruin, and S. Gallagher, 41–70. Oxford University Press.

Rietveld, E. and J. Kiverstein. 2014. "A Rich Landscape of Affordances." *Ecological Psychology* 26 (4): 325–52.

Saks, E. 2007. *The Center Cannot Hold: My Journey Through Madness*. Hachette Books.

Sass, L. 2004. "Affectivity in Schizophrenia: A Phenomenological View." *Journal of Consciousness Studies* 11 (10–11): 127–47.

Segundo-Ortin, M. 2020. "Agency from a Radical Embodied Standpoint: An Ecological-Enactive Proposal." *Frontiers in Psychology* 11: 1319.

Stanghellini, G. 2001. "A Dialectical Conception of Autism." *Philosophy, Psychiatry, & Psychology* 8 (4): 295–98.

Stern, D. 2010. *Forms of Vitality: Exploring Dynamic Experience in Psychology, the Arts, Psychotherapy, and Development*. Oxford University Press.

Szasz, T. 1960. "The Myth of Mental Illness.' *American Psychologist* 15: 113–18.

Thompson, E. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of the Mind*. Belknap Press.

———. 2011. "Reply to Commentaries." *Journal of Consciousness Studies* 18 (5–6): 176–223.

Trémeau, F. 2006. "A Review of Emotion Deficits in Schizophrenia." *Dialogues in Clinical Neuroscience* 8 (1): 59–70.

van Dijk, L. and E. Rietveld. 2017. "Foregrounding Sociomaterial Practice in our Understanding of Affordances: The Skilled Intentionality Framework." *Frontiers in Psychology* 7: 1969.

Vermeulen, P. 2015. "Context Blindness in Autism Spectrum Disorder: Not Using the Forest to See the Trees as Trees." *Focus on Autism and Other Developmental Disabilities* 30 (3): 182–92.

Weber, A. and F. J. Varela. 2002. "Life after Kant: Natural Purposes and the Autopoietic Foundations of Biological Individuality." *Phenomenology and the Cognitive Sciences* 1 (2): 97–125.

Weichold, M. 2018. "Situated Agency: Towards an Affordance-Based, Sensorimotor Theory of Action." *Phenomenology and the Cognitive Sciences* 17 (4): 761–85.

Yamaguchi, S. and K. Onoda. 2012. "Interaction between Emotion and Attention Systems." *Frontiers in Neuroscience* 6: 139.

Yirmiya, N., C. Kasari, M. Sigman, and P. Mundy. 1989. "Facial Expressions of Affect in Autistic, Mentally Retarded and Normal Children." *Journal of Child Psychology and Psychiatry* 30 (5): 725–35.

Received: 15 August 2022
Accepted: 11 January 2023
Published: 15 June 2023

Grief over Non-Death Losses: A Phenomenological Perspective

Matthew Ratcliffe and Louise Richardson

Abstract

Grief is often thought of as an emotional response to the death of someone we love. However, the term "grief" is also used when referring to losses of various other kinds, as with grief over illness, injury or impairment, unemployment, a relationship breakup, or loss of significant personal possessions. In this paper, we address the question of what, if anything, the relevant experiences have in common. We argue that grief over a bereavement and other experiences of loss share a common phenomenological structure: one experiences the loss of certain *possibilities* that were integral – and perhaps even central – to the unfolding structure of one's life. Grief can thus be conceived of in a broad but still univocal way. To develop this position, we focus on the example of grief over involuntary childlessness, where the lack of a concrete, historical object of emotion serves to make explicit the way in which grief concerns future possibilities. We go on to suggest that the phenomenological complexity, diversity, and prevalence of grief are obscured when approached via an abstract, simplified conception of bereavement.

Keywords: bereavement; childlessness; grief; identity; life structure; loss; possibilities

1. Introduction

Grief is often taken to be an emotional response to the death of someone we love. However, the term "grief" is also used in a broader way that encompasses losses of various other kinds. Grief, or at least grief-like experiences, are said to arise in response to a range of situations and events, including relationship breakups, serious illness or injury, loss of employment, homelessness, and other circumstances (Harris, 2020). It should not simply be assumed that the English language term "grief" refers to a singular phenomenon; perhaps some of the relevant experiences have little in common. However, it has been suggested that there are indeed important similarities between experiences of bereavement and other losses. Given this, there have been calls for a new research field, focused around a more inclusive conception of loss (Harvey and Miller, 1998). Nevertheless, there remains the concern that a permissive conception of grief might bring together a diverse assortment of life circumstances and emotional experiences that have little in common. For instance, how – if at all – are we to distinguish grief from a chronic sense of unfulfillment, where certain achievements appear forever beyond one's reach? And what of regret, disappointment, failure, guilt, shame, sorrow, and loneliness? It will not suffice to insist that an experience of grief is somehow qualitatively distinct from all of these. For

one thing, any such claim would need to be substantiated. Furthermore, grief is not an episodic emotion but, more plausibly, a temporally extended process that varies markedly over its course and incorporates emotional experiences of many other kinds (Goldie, 2012; Ratcliffe, 2017a, 2022). Hence, we are concerned with the question of what – if anything – distinguishes this temporally extended pattern *as a whole* from other types of episodic and diachronic emotional experience. Closely related to this is the issue of what the *objects* of grief share in common: which characteristics make something an appropriate object of grief, rather than – say - regret or dismay?

In what follows, we will set out a phenomenological case for a broad conception of grief, which addresses concerns about heterogeneity by identifying something that is common to experiences of loss and also distinguishes them from other forms of emotional experience.[1] We do not thereby seek to preclude an additional, bereavement-specific conception of grief. Indeed, it is plausible that experiences of personal loss are distinctive in important ways (Ratcliffe, 2020, 2022). Rather, our position is that it can *also* be informative to think of grief in broader terms, given the structural similarities between various experiences of loss. Furthermore, by doing so, we can enrich our understanding of bereavement grief.

Our discussion will focus specifically on the example of grief over involuntary childlessness. We have chosen this example because it is particularly effective in showing how grief is not simply *past-directed*, but centrally concerned with losses of *future possibilities*. What also becomes apparent is that these possibilities are integral to *who* one is – to one's life structure, sense of identity, or experiential world. Of course, it should not simply be assumed from the outset that grief over childlessness is informatively similar to other experiences that might be referred to as grief. However, we will go on to argue that this same analysis captures what is common to and distinctive of grief experiences more generally, including grief over a death.[2] Importantly, it also enables us to recognise the phenomenological complexity of loss. This complexity is obscured when the paradigm case of grief is taken to be a decontextualised scenario where one person dies swiftly, and perhaps unexpectedly, after which another person grieves over their death. In bereavement and other situations, grief is seldom so neatly bounded. This is largely because its *object* is not neatly bounded, rendering the experience difficult to conceptualise and articulate.

2. Grief Over Involuntary Childlessness

For a broad conception of grief to be tenable, a sufficiently encompassing account of grief's object is required. The object of grief – what it is *about* – is often taken to be the event of someone's death or their being dead (Moller, 2007; Marušić, 2018). This restricts grief to the context of bereavement. But it also renders grief anomalous, given that emotions of other kinds are not generally classified by appealing to their concrete objects. For instance, one could – in principle – be afraid of anything (although whether an entity, event, or situation is an *appropriate* object of fear depends on the particulars of one's situation). We thus identify *types*

1 We do not go so far as to maintain that our phenomenological account maps neatly onto how the English language term "grief" operates in everyday discourse. Talk of grief may well play a number of different roles in describing, conveying, and expressing emotion. For instance, referring to one's experience as one of "grief" could serve to emphasize the level of difficulty and distress involved, rather than being an attempt to categorize the experience accurately.

2 This account will also accommodate grief over the death of nonhuman animals, but does not apply so obviously to the grief *of* nonhuman animals. Regardless of whether other animals experience something that might be termed "grief", the experiences we are concerned with require a temporally organised life, involving projects, commitments, expectations, and aspirations that may stretch far into the future. Human lives are distinctive in this respect. For the same reason, human infants may not experience loss in quite the way described here.

of emotion, such as fear, relief, anger, and happiness, rather than the likes of shark-fear, not-losing-one's-job-relief, anger-at-queue-jumpers, and woodland-walk-happiness. To render grief consistent with wider practice, we could appeal instead to its distinctive *formal* object: a value property possessed by its concrete object, which is associated with an emotion of that type (e.g., de Sousa, 1987; Teroni, 2007). This would be in line with how other types of emotion are identified by their characteristic formal objects, as with the case of fear where the formal object is threat. A natural candidate for the formal object of grief is loss. To make this specific to bereavement, we could refer more specifically to the loss of a *person*. That would still accommodate the likes of relationship breakups and close family members moving overseas. So, we could refer still more specifically to *irrevocable* loss. Although this would not suffice to identify an emotion experienced exclusively in response to the deaths of those we love, the formal object of grief would approximate this very closely.

We have no objection to such an approach, but the question remains as to whether there is also a broader type of grief experience, the formal object of which is *loss* of a more general kind. The answer depends on whether or not this alleged "grief" can be distinguished from a host of other emotional responses, such as disappointment, regret, remorse, frustration, resignation, sorrow, and guilt. We will suggest that it can, while adding that the relevant phenomenology, which it shares with emotional responses to bereavement, has been obscured by tendencies to (a) construe grief as a reaction to *historical* loss, and (b) think of loss in ways that are too generic and abstract. To make explicit and challenge both tendencies, we will focus on the example of grief over involuntary childlessness, which can involve a sense of loss that is not concerned primarily with something specific, concrete, and past. To do so, we will turn to some first-person accounts obtained via a qualitative survey of bereavement grief. Participants were invited to provide open-ended, free-text responses to a series of questions about their current or past experiences of grief. In total, 265 responses were received, most of which included detailed descriptions of emotional experience. However, to our surprise, 29 respondents did not address bereavement grief, but instead grief over childlessness. Some of these respondents remarked on a tendency to overemphasise bereavement when thinking about grief, and on a consequent failure to recognise and acknowledge experiences such as theirs.[3] These responses were all submitted by women, who had been directed to our survey by the network Gateway Women, which provides support for women who are involuntarily childless.[4] However, similar experiences of loss are also described by some childless men (Hadley and Hanley 2011, 61; Hadley, 2021).

It is not always clear when an experience involves grief *over* childlessness, rather than grief over bereavements or losses of other kinds. Some respondents who mentioned childlessness also described feelings of loss associated with specific historical events or sequences of events, including abortions, miscarriages, failed IVF treatments, illnesses and treatment outcomes, relationship breakups, and the deaths of relatives. However, in the 29 testimonies we refer to here, the primary focus was on not having children. Respondents explicitly identified the principal object of their grief as childlessness *per se*, even when they also described more specific, historical losses. Involuntary childlessness was attributed to a variety of causes and concerned – for the most part – not having children at all, rather than a more specific inability to have biological children. Some remarked that they found the frequent question "why not adopt?" insensitive, in failing to acknowledge the emotional and practical complexities of the situation. As one respondent wrote, "The phrase 'why not just adopt' is the bête noire of all childless women. It is said [as if it is like] going to the supermarket. [….] Everyone wants to suggest a fix, but there is no easy fix" (#264).

3 The study was conducted from 2020 to 2021, as part of the AHRC-funded project "Grief: A Study of Human Emotional Experience". For further details, see Ratcliffe (2022).

4 The network founder, Jody Day, has also published a detailed account of her own grief over childlessness, which emphasizes distress caused by predominant social and cultural attitudes (Day, 2016).

In what follows, we are concerned with childlessness as an object of grief. At the same time, we acknowledge that the relevant circumstances are diverse and that emotional experiences of loss further involve various different concrete objects of emotion. One might take the line that grief over childlessness is not, strictly speaking, an experience of *loss* at all, as nothing concrete was *taken* from the person. After all, there is surely an important difference between emotional experiences of losing something and of never having had it in the first place. However, we will instead show how first-person accounts of grief over childlessness challenge that way of thinking about loss.

Some respondents to our survey emphasize just how emotionally painful involuntary childlessness can be, comparable even to the effects of bereavement: "The grief over a person that someone has welcomed/wished for/loved in advance but was never there can be as devastating as the grief over the death of a person that has lived a real life" (#213). Of course, that an emotional experience affects someone so profoundly that they compare it to bereavement does not imply that the two experiences are also similar in *kind*. It could be that emotional responses to childlessness – however distressing they might be – are quite different from experiences of actually losing someone or something important. Perhaps they involve painful feelings of disappointment, regret, sorrow and longing, which do not add up to grief. However, although these survey respondents do not attribute their grief to the effects of particular historical events, they still describe it in terms of *loss*:

> Because I haven't actually lost a person but lost the life I thought I would have, which was children, it feels all-consuming. (#228)
>
> This is the death of a future life. (#222)
>
> It is the loss of a dreamed-of future, a life you have imagined since you were a child. (#264)
>
> I had to say goodbye to the child I desperately wanted but was never able to have, they died in my heart and will never leave. (#258)
>
> I am experiencing grief and loss around being unable to have children. I am grieving the future children I imagined and believed I would have but am unable to. (#265)

What seems to be lost in these cases is a significant *possibility* or cluster of interrelated possibilities that was sought, anticipated, and often imagined in varying degrees of detail. Now, the failure of something highly desirable to occur is not ordinarily a cause of grief. For instance, not winning the lottery ordinarily brings only mild disappointment, if that. An important difference, though, is that having a child was not just taken to be possible – it was also anticipated, perhaps with such confidence that the prospect of childlessness was never really contemplated. Furthermore, having children was something that these women cared about deeply and actively sought. Nevertheless, it remains doubtful that anticipating, caring about, and investing considerable energy into something makes its non-occurrence a sufficient cause for grief or an appropriate object of grief. Instead, it might elicit some combination of surprise, disappointment, sadness, frustration, regret, disillusion, and demoralisation. Take the scenario – familiar to many academics – of having a major grant application rejected. This can bring marked and prolonged disappointment, spanning a range of emotions, but it is not clear that some or all of these are sufficient for grief. A conception of grief that accommodates all major disappointments seems too permissive, failing to capture a more specific type of emotional experience. A failed grant application is not typically experienced as the *loss* of something (at least not in the relevant

sense of "loss"), but as something that was never gained, however deserving one may have been.[5] What, then, marks the difference between experiences like this and those that we might term "grief"?

What is crucial is that the lost possibilities towards which grief is directed are not just much-desired states of affairs in an imagined or anticipated future; these possibilities are also integral to a sense of *who one is now* or – at least – to who one *was* before full acknowledgement of their foreclosure. Consistent with this, experiences of lost possibilities associated with involuntary childlessness are sometimes described in terms of the loss, diminution, disruption, or transformation of a kind of *identity*:

> I live with the grief for the children I never had and the identity I lost as a result. . . . Sometimes it was stabbing grief with sobbing, other times it was a numbness, an incomprehension of how I had arrived at that point, who would I be now? (#226)
>
> This grief was the worst ever. It was the loss of my dreams and future. The loss of who I was meant to be. (#201)
>
> I am a completely different person. (#199)
>
> The nature of my relationship was with an identity that I would be a mother and then repeatedly to have that dashed and eventually destroyed because of my age. (#241)

Thus, appreciation of the inability to become someone or something affects who one is now. An anticipated or imagined future self or future situation was integral to one's current sense of self. With the acceptance that certain outcomes are irrevocably beyond reach and are now counterfactual rather than futural, one's identity is profoundly altered. This disrupted sense of *identity* can equally be described in terms of a *life structure* or an *experiential world*; one's life is turned upside down or on its head, and one comes to inhabit a different world (Ratcliffe, 2017a, 2017b, 2019, 2020, 2022). It is this, we suggest, that distinguishes the experience from one of disappointment. Where the latter is concerned, a future possibility might be negated or foreclosed, but it is not experienced as lost in quite the same way, given that the anticipated state of affairs was not integrated into one's life to the same extent. Hence, although a failed grant application is not ordinarily an object of grief, one might well experience grief in a scenario where repeated disappointments eventually lead to the loss or abandonment of a career in which one was heavily invested. In the case of grief over childlessness, one can similarly be tasked with reorganizing a life. Survey respondents were at different points in this process: "My identity has shifted gradually. . . . which brings ease" (#226); "loss of one's identity as a mother and becoming a mother is endless" (#241).[6]

Grief is thus distinctive insofar as it involves recognising and comprehending the implications of lost possibilities for the structure of one's life or world, and – over time – coming to reorganise one's life accordingly (Attig, 2011; Ratcliffe, 2017a, 2017b, 2019, 2020, 2022; Fuchs, 2018; Read, 2018; Mehmel, 2021; Ratcliffe and

5 This is not to say that a life could not be organized in such a way that a failed grant application would be experienced as a loss. Rather, the point is that this is typically not the case.

6 In an interesting recent discussion, Mehmel (2021) appeals to a similar example: grief over a stillbirth, where there is no history of shared habits and projects involving the deceased. Drawing on this example, Mehmel also observes that grief is not simply past-directed, but oriented towards future possibilities of the self. However, it remains the case here that one's grief is concerned with a concrete, historical loss. Grief directed at childlessness per se is thus different; it does not have a concrete, historical object. Hence, the example works differently, pushing us towards the conclusion that the object of grief is a loss of possibilities.

Byrne, 2022). In referring to an experiential world, life structure, or sense of identity, what we have in mind is a network of interrelated projects, commitments, relationships, and expectations, which could be loosely termed a person's distinctive "value system". This constitutes the unique organisation of a life, an answer to the question of *who* someone is: I am a teacher; I am a husband and parent; I am committed to this and that; I enjoy doing those things; these are my habitual, daily activities; these are my future plans (which reflect my projects and commitments); this is how the current organisation of my life is shaped by those plans.

A life structure is thus closely related to what Korsgaard (1996) calls a "practical identity": a set of categories through which we think about ourselves and engage with the social world.[7] Practical identity – in contrast to some bare, undifferentiated sense of being a singular locus of experience – includes the likes of being a parent, spouse, or teacher, along with one's sex, gender, sexuality, political and religious affiliations, and numerous other categories that incorporate action-guiding norms and shared expectations. When someone describes a loss or change of identity in realising that they will not have children, "future parent" was one of the categories that made up their identity. Another relevant notion is that of the "narrative self", conceived of as a unifying, temporally organised story that comprises *who* one is, something that need not be explicitly *told* in order to constitute one's distinctiveness and coherence.[8]

However, when we say that the loss experienced in grief impacts upon *who* one is, what we have in mind is broader in scope than both narrative and practical identity. Values – broadly construed – can be reflected in emotional experiences of situations and events without their being explicitly or implicitly integrated into an organised narrative. Furthermore, not all aspects of a life structure are conceptual or articulable. Much of the relevant organisation is integral to an experiential world that our more specifically directed experiences, thoughts, and activities take for granted. Things appear immediately significant to us in ways that reflect an established value system – they *matter* to us as they do in light of our established projects, commitments, relationships, and pastimes. For the most part, how things matter is grasped pre-reflectively and habitually, as inherent in our surroundings. The delayed train time on the departure board, the football that bounces off the goal post, the child falling off their bike, and the five hundred new messages in one's inbox appear immediately significant in distinctive ways that reflect one's values. Whether or not we regard such experiences as specifically *perceptual*, it is plausible to maintain that gauging how something matters to us does not ordinarily require explicit inference from experience. Nor should it be construed as specifically *cognitive*, at least not in a way that is to be contrasted with what is bodily, habitual, and felt. It is – in part, at least – through a range of felt, bodily dispositions that we experience and respond to significant events (Ratcliffe, 2008, 2015, 2022). To classify all of this as narrative would be to adopt a conception of narrative that is too broad to be informative. A similar point applies to practical identity, construed as a "description under which you value yourself" (Korsgaard, 1996, 101). Much that makes us who we are would not figure in such a description, if by this is meant something we can articulate. In fact, it may only become salient to us, and perhaps to others as well, with the loss of certain possibilities that were previously taken for granted.[9]

7 Cholbi (2022) also highlights the relevance of practical identity to grief, although specifically for understanding which deaths we grieve over. On his view, "the more central another person is to our practical identity, the greater cause we have for grieving them upon their deaths" (31).

8 See, for example, Zahavi (2007) for a good discussion of conceptions of the "narrative self" and for the view that they do not add up to a comprehensive account of self.

9 Thus, as Cholbi (2022) suggests, there may be interesting relationships between experiences of grief and the acquisition of self-knowledge. Experiences of loss have the potential to make salient, disrupt, and problematise aspects of oneself that might otherwise be taken for granted.

Importantly, the structure of a life is not something that ordinarily remains constant, something that is disrupted only by occasional life events. Rather, it is dynamic in nature and also experienced as such. As indicated by descriptions of grief over childlessness, who one is *now* depends on the possibilities of becoming someone or something. One's life is oriented around those possibilities, and the significance of unfolding events is experienced as a movement towards or away from them.[10] With that movement, the possibilities around which a life is organised also change – often in a subtle, gradual fashion. However, where an event or situation has profound implications, the temporally extended process of recognition, comprehension, and adaptation can involve profound upheaval. Faced with lost possibilities, such as that of becoming a parent, one may experience "the death of an assumed way of life" (#253) or the "collapse" of the "world as I knew it" (#258), something that involves a sense of uncertainty over how to go on (Ratcliffe, 2017b, 2020; Mehmel, 2021).[11] A scenario where one actually had the winning lottery ticket in one's pocket but unintentionally discarded it is importantly different.[12] There might be a profound feeling of regret or disappointment, involving a distressing and lingering sense of what could have been. Nevertheless, this does not amount to grief, at least when the relevant possibilities were not integral to the shape of one's unfolding life.[13]

3. Lost Possibilities

For current purposes, the example of grief over involuntary childlessness is informative in showing how a profound sense of loss can be experienced without there being a historic, concrete object of emotion. One might think that this makes it an exception to the rule. However, we will now suggest instead that the example serves to make salient something that applies more generally. In those instances of grief (including bereavement grief) where there *is* a concrete object of emotion, the experience is similarly concerned with a loss of life possibilities. It can be added that something is an *appropriate* object of grief insofar as it constitutes such a loss. This point can be couched in both phenomenological and nonphenomenological terms. There is a fact of the matter concerning the potential and actual implications of an event or situation for a person's life. For instance, where project P is central to one's life and the integrity of P presupposes R, the irrevocable absence of R amounts to a loss regardless of whether it is experienced as such. However, we are concerned more specifically with the associated phenomenology, which involves experiencing and navigating – sometimes over a lengthy period of time – the negation of possibilities that were integral to one's life.

10 Consistent with these observations, Di Paolo (2020, 2021) suggests that emotion is integral to the experience of becoming. Recognising the significance of something does not involve the rupturing of a previously static and complete sense of self. Instead, he suggests, we are "unfinished creatures" and "to feel an emotion is to experience a moment in our own becoming" (2020, 229).

11 Appealing to themes in the work of Merleau-Ponty, Ratcliffe (2020, 2022) refers to this sense of uncertainty as an experience of pervasive "indeterminacy". Mehmel (2021), drawing on the work of Harbin (2016), refers to the same phenomenon as "disorientation".

12 As Jody Day (2006, 220) writes, "the shift in identity from being a woman who hopes, one day, to become a mother to one who knows, without question, that it's never going to happen is so huge that it throws everything into question". This is not to say that any subsequent identity must be organised around one's childlessness. As Day (2006, 5) also observes, accommodating something does not require making it central to who one later becomes.

13 That is not to deny that one could, under certain circumstances, experience genuine grief over not winning the lottery. Suppose one's entire life had been organised around doing so. In eventually resigning oneself to the fact that it will not happen, one might then experience a significant loss of life possibilities. Here, the grief would be genuine and in some sense appropriate, in that it would accurately reflects the implications of a situation for one's life structure. Grief in such circumstances can nevertheless be inappropriate in the sense that relates to the moral or pragmatic impropriety of one's pursuits, including the overall organisation of one's life. The same could be said of numerous far-fetched scenarios, such as a life that is organised around the prospect of becoming Emperor of the Galaxy, which is eventually renounced. See D'Arms and Jacobson (2000) for a discussion of the significance of the distinction between the two ways in which an emotion can be appropriate. Unless otherwise stated, we use "appropriateness" here in the former sense of accuracy or "fittingness".

To suggest that grief in general is concerned with the loss of possibilities is not to suggest that it lacks a concrete object. Losses of possibilities can also be described in more concrete terms. Nevertheless, even where grief is associated with a particular event, it would be an oversimplification to say that it has a single, concrete object. In the case of bereavement, grief involves intermittent preoccupation with many things: the event of the death, how it occurred, the fact that someone is dead, what I have lost, what that person has lost, what might have been, how I will cope without them, what the future holds, the impact upon others whom I care about, and so forth. All of these objects of emotion can be understood as integral to a wider-ranging loss of possibilities.[14] Furthermore, there is no straightforward distinction to be drawn between the formal and concrete objects of grief. Instead, losses of possibilities are experienced with varying degrees of concreteness (Ratcliffe, Richardson, and Millar, 2022). What is lost may appear quite precise or utterly inchoate, with the content of one's emotional experience varying over time. Likewise, grief over childlessness might concern being a parent, having a child, having a girl or boy, having a child with a certain name, loving that child, or even being a grandparent.

What we have not yet shown is that grief over a death and grief over childlessness do indeed have a common phenomenological structure. It turns out, though, that there are good reasons for thinking that grief in response to a death, like grief over childlessness, is not simply about something that has happened – both concern losses of *future* possibilities. In support of that position, consider the differences between our emotional responses to bereavements and our responses to certain substantial changes in people and relationships (especially when those changes occur over extended periods of time). The parent of a teenager does not usually grieve over the loss of the baby, toddler, or young child who once was, even though all of the distinguishing properties of the child may have changed, and even though the relationship is now radically different to how it once was. Granted, there may be moments of nostalgia, sorrow, and regret. Nevertheless, the combination of these is altogether different from profound grief over a death. How is this to be explained? The contrast, we suggest, stems from an intuitive sense of the difference between possibilities being actualised and built upon, as when a child grows from baby to teen, and possibilities being negated, as when someone dies. As the child grows older, the process is – in most instances – experienced largely as one of unfolding development rather than loss. Hence, to appreciate what is lost – and also experienced as lost – when a person dies, we need to acknowledge that it is not something concrete and historic. Instead, it is a certain kind of *potentiality*.

One might object that this analysis fails to accommodate instances of bereavement where we grieve despite regarding the death itself in a mostly positive light—perhaps an end to someone's suffering in old age, at a time of life when they had accomplished all they set out to do. Why would we experience the loss of something future-directed in such a case? However, a range of different circumstances can be accommodated via the simple acknowledgement that, in bereavement and other situations of loss, the balance between loss of *my* possibilities, *their* possibilities, and *our* possibilities can vary markedly. For instance, when faced with unemployment, one's first thoughts might concern oneself or, alternatively, the impact on one's family. In this respect, bereavement is plausibly distinctive: we are confronted with the complete, irrevocable loss of someone else's possibilities or, at least, those possibilities that depended in some way upon that person's

14 In the case of bereavement, another candidate object of grief is the loss or radical transformation of an interpersonal relationship (Cholbi, 2019). As a concern with relationship loss also involves attending to the fact that someone has died, we might say that the object of grief is broader in scope than a person's death or their being dead. However, we suggest that this is still overly restrictive – grief involves a wider-ranging engagement with losses of possibilities.

continuing experience and agency.[15] However, bereavement experiences involve various different balances between possibilities that are experienced as mine, theirs, and ours. Grief over the death of a suffering and accomplished loved one may be concerned with *my* and *our* lost possibilities, more so than *theirs*.

In returning to the case of grief over involuntary childlessness, it might seem that the balance here is skewed towards one's own possibilities. The sense of loss could also envelop the children that a *partner* will never have or that *we* will never have together. Nevertheless, it might be assumed that the absence of children is experienced as one's own loss or our loss, rather than a loss that *they* incur by not coming into being. As one respondent reflected, "maybe I don't feel their absence as much as I feel what I am not" (#214). But the children one will not have are sometimes described with remarkable specificity – they might be a girl or boy, with a name and certain characteristics. Some respondents go so far as to ascribe a kind of past or ongoing existence to children who never came into being:[16]

> Part of my healing has been to bring her to life. I say her name out loud now, I can see her hair, her smile, I talk about her to the one person who allows me to talk about her. It warms my heart and gives me great peace to acknowledge her existence. [....] I had a bracelet made with her name engraved in it so I can wear it whenever I need to feel her near me. (#212)

> It is hard to explain, but I can feel their love and my love for them when I am bringing them up in my mind and heart. (#223)

> Although they never came to this world, I feel and know them as real, truly existing persons that I never will have the chance to get to know. (#213)

Testimonies such as these are challenging to interpret. For one thing, it is not clear whether, when, or how to distinguish one's emotional experience of a child who never came into being from the imaginative embellishment of a less determinate loss experience (which could involve borrowing from narratives and practices associated with bereavement). However, they do suggest that grief over childlessness may not be limited to the lost possibilities of the living. Insofar as one is able to contemplate an absent child as an individual who was denied being, one's grief might extend to their possibilities as well.

In other circumstances too, relationships between one's own possibilities and those of others can be complicated, and are sometimes riddled with tensions. Take the case of a parent suffering from "empty nest syndrome" after their adult children have left home. It is arguable that this can approximate an experience of grief, even though it involves the lives of one's children progressing in ways that were expected and viewed positively. One single father describes the predicament as follows:

15 That said, it could be maintained that certain possibilities are able to remain and be realised posthumously. For instance, it is arguable that a person's desires can still be satisfied or frustrated after death (See Solomon, 2004, for further discussion). There is also the difficult question of how various beliefs concerning life after death shape our experiences of loss. Even so, it seems plausible to maintain that death affects a person's possibilities in a manner that is more profound than all, or almost all, events that occur in life.

16 This points to a potentially interesting extension of "continuing bonds" approaches to grief, which propose that bereavement involves revising and sustaining a relationship with the person who has died, rather than "letting go" (Klass, Silverman, and Nickman, 1996). Certain descriptions of grief over childlessness suggest that it may also be possible to form, sustain, and develop a sense of interpersonal connection with someone who never was.

> When they left I lost direction. . . . I found it quite difficult to know what my role was. . . . As a parent, from the minute they go to playgroup their relationships impact on you. When they moved out all that suddenly went and I missed it. Suddenly I hadn't got anybody to talk to me about their friends and stuff. So there was an emptiness. (Dodd 2011, 72)

In circumstances like these, *my* loss of possibilities can at the same time be *your* actualization and further pursuit of possibilities. So, the sense of loss is equivocal. Certain of my own possibilities are negated, while others – which include striving to advance someone else's possibilities – have been actualised. In this way, a grief-like experience can co-exist with a sense of achievement, gratitude, and even joy. There are many other scenarios where the possibilities central to one's own life include striving to actualise someone else's possibilities, in ways that involves losing - in the process - other possibilities that are integral to who one is. Human relationships in general can be thought of in these terms, as fraught with tensions between development and loss. Hence, by attending to grief as a loss of possibilities, instead of fixating upon a singular, clearly defined, historical object of grief, we can come to better appreciate the complexities of emotional experience, including tensions between possibilities lost and possibilities realised.

To further complicate the picture, lost possibilities also relate in various different ways to one's biography and to time. For childless men interviewed by Hadley (2021, 4), "fatherhood was viewed as a reconnection, repayment, repeat or replacement of their childhood experience". As this indicates, the emphasis that we have placed upon possibilities should not be taken to imply that grief is exclusively future-oriented. Possibilities, as experienced pre-reflectively or contemplated explicitly, can have a complex temporal structure. For example, one survey respondent described "a loss of memories that would never happen" (#21). Here, what is lost is not just the possibility of certain events occurring in the future, but also that of remembering them at a later date. As this suggests, lost possibilities should not simply be conceived of as future-directed in contrast to past-directed. Instead, the two are inextricable. How past events *matter* to us reflects where we are heading; it depends on which possibilities remain open and which have been extinguished. So, even where grief does not have a concrete, historical object, it retains a concern with the past. Complementing this observation, Goldie (2012, chap. 3) draws attention to the ironic structure of memory in grief. One continues to experience the significance that events had when they occurred and for a period thereafter. But one's recollections also integrate what is now known: the person has since died and the possibilities towards which past events point, the paths to which they lead, have therefore shifted radically. These two perspectives intermingle, although uneasily. Sartre ([1943] 1989, 498-9) makes the broader point that there is a sense in which we *choose* the past. How past events matter to us – how they feature in significant, temporally organised patterns – depends on which future possibilities are currently central to our lives, which possibilities we make our own. So, a loss of future possibilities is also a shift in the significance of the past. In the case of involuntary childlessness, there may not be a single, most salient past event or pattern of events. Nevertheless, what it shares with other experiences of loss, including bereavement, is a wider-ranging significance attaching to events in one's biography, in light of life possibilities and their negation.[17]

17 Here, we agree with Mehmel (2021) that the phenomenology of grief cannot be understood solely in terms of a loss of previously established, shared practices, and that it should also be conceived of in terms of one's orientation towards the future. See also Ratcliffe (2017a, 2017b, 2019, 2020, 2022) for such a position. However, it is important to distinguish this clearly and consistently from the stronger claim that grief concerns our sense of what the future holds instead of the habitual organisation of experience, thought, and activity. That profound grief need not always involve the loss of a life structure that was shared with the deceased remains consistent with grief's involving the disturbance of one's previously established sense of what is and what was. For instance, a loss of possibilities that were central to the overall organisation of one's life is at the same time a transformation of the significance of past events and how they hang together in one's biography, even where the loss in question is not attributed to specific historical events. We thus take it that, in engaging with a loss of possibilities, one's experiences of past, present, and future are inextricable.

Hence, the common theme is that grief involves engaging – over a period of time – with how the shape of one's life has been, continues to be, and will be affected by losing certain possibilities. This generalises to experiences of loss in a wide range of other situations. For example, consider the following passage from a book by the clinician Paul Kalanithi, which describes the moment when he was about to receive a terminal diagnosis:

> A young nurse, one I hadn't met, poked her head in.
>
> "The doctor will be in soon."
>
> And with that, the future I had imagined, the one just about to be realized, the culmination of decades of striving, evaporated. (Kalanithi, 2017, 16)

Confirmation of his illness is experienced *as* an abrupt and conspicuous loss of possibilities, as the negation of a future around which his life has been organised. Later, there is a more determinate appreciation of all those possibilities that will no longer be actualised, a growing realisation that amounts to the loss of an identity he had taken for granted:

> My life had been building potential, potential that would now go unrealized. I had planned to do so much, and I had come so close. I was physically debilitated, my imagined future and my personal identity collapsed, and I faced the same existential quandaries as my patients faced. (Kalanithi, 2017, 120)

What distinguishes the experience of grief from other forms of emotional experience is not a specific feeling quality, evaluation, or other ingredient that is present at a specific time. There is no single moment at which we experience a loss of possibilities and no single way in which we do so. We experience and respond to more profound forms of loss over lengthy periods of time, in ways that encompass habitual patterns of thought, ways of experiencing our surroundings, autobiographical memory, bodily anticipation, and more specific emotional experiences of various different kinds. A grief process can involve immediate, pre-reflective experiences of absence and lack, sliding into habitual patterns of thought and then realising that they no longer apply, having a sense of detachment from one's surroundings (which no longer embody significant possibilities for projects and other activities in ways they once did), feeling disconnected from other people (who carry on regardless), and being struck by tensions between propositional beliefs that reflect the realities of one's situation and an experiential world that seems to endure – at least for a time – in spite of them (Ratcliffe, 2022). All of this becomes more apparent once we attend to the dynamic ways in which networks of possibilities and their negation are experienced, sometimes spanning all aspects of a human life – past, present, and anticipated. Taking a case where there is no single, discrete, concrete object of emotion as our exemplar prompts us to do so.

4. Bereavement Grief Revisited

We have argued that there is a broad kind of loss experience with a distinctive phenomenological structure, which encompasses bereavement grief and also a wide range of non-death losses. In this section, we will elaborate on our suggestion that conceiving of grief in this broad way stands to enrich our understanding of bereavement grief, enabling us to better appreciate its complexity and diversity. In the case of grief over involuntary childlessness, we find an experience without a circumscribed historical object or a clear-cut

starting point. One might suggest that this is to be contrasted with bereavement: "it's a long slow grief that comes and goes; it's not like the sudden death of a loved one" (#206). However, it is important to acknowledge that experiences of bereavement vary considerably, depending in part on the circumstances surrounding a death. If we think of a death as something that happens swiftly, perhaps unexpectedly, and consider a person's subsequent grief in abstraction from their situation as a whole, it might well seem that bereavement grief is directed at a specific, clearly demarcated event or situation and that it is quite different from the sense of loss sometimes associated with childlessness. But deaths can be expected for some time and are often preceded by other losses (such as losses of bodily or cognitive abilities). Many possibilities – for *them*, for *oneself*, and for *us* – may be lost before the person dies. So, in seeking to develop an adequate account of how people experience loss, even in the context of bereavement, it is important not to place too much weight upon a generic, simplified construal of bereavement. It can be helpful to begin instead with a more encompassing and less tidy conception of grief and its objects.

The example of childlessness grief also draws attention to a range of other factors that complicate the structure of grief experiences, in ways that sometimes render them difficult to conceptualise and convey to others. One useful concept here is that of "ambiguous loss" (Boss, 1999). This concerns predicaments of two broad kinds that can render one's experience of loss equivocal or conflicted. Sometimes, one does not *know* exactly what has happened to a lost object. A body might never have been recovered, the circumstances of a death may remain unknown, or perhaps the person has simply disappeared. In the other kind of case, it could be said that the situation is *metaphysically* rather than *epistemically* challenging. For instance, when a person has changed radically, we might wonder whether and in what sense it is *still them*. The common theme is that we are suspended indefinitely between two scenarios, one where a possibility is negated and the other where it remains. Faced with this, we may vacillate between hope and resignation. Both forms of ambiguous loss appear in first-person accounts of involuntary childlessness. There might be no specific moment at which it becomes fully clear that one will not have children; one hovers between possibilities for an extended period of time. In addition, the nature of one's loss can be challenging to make sense of – what, exactly, has been lost, when did the loss occur, and by whom was it incurred?

Another complicating factor is whether or not there are established norms, rituals, and narratives for responding to one's own and others' experiences of loss. In the case of childlessness, there is no grave, no anniversary, no recorded name, no special place, no memorial service, and no shared practices for narrating and making sense of the loss.[18] Partly for that reason, grief over childlessness is sometimes described in terms of "disenfranchised grief", a term introduced by Kenneth Doka to capture a variety of ways in which grief can be unacknowledged, misunderstood, or denied legitimacy (Doka, 1999, 2002).[19] First-person accounts of involuntary childlessness identify several factors that contribute to disenfranchisement, including a lack of shared practices:

> The deaths of both my grandfathers in 2015 and 2020 were honoured through rituals and funeral day events. [....] Their death is absolute whereas my losses are both not only not seen but also, they sweetly haunt me through life-long grief. (#229)

18 However, this has started to change. For example, the Mariposa Trust welcomes women who have not been able to have children to its "Saying Goodbye" ceremonies.

19 See also Day (2016) for a detailed discussion of the various causes of disenfranchisement among those who grieve over involuntary childlessness.

Some also mention, as a source of disenfranchisement, living in a society that is structured around parenthood, where much that people do relates in one or another way to having children. There is thus a failure to accommodate those who are involuntarily childless and to recognise the extent of their loss: "Society does not make room for those women who are childless not by choice" (#224). Also contributing to disenfranchisement is the tendency to think of grief and loss in exactly the sort of way that we have challenged here – as a matter of responding emotionally to the historic subtraction of something or someone concrete. Because grief is conceived in that way, it is not even acknowledged that there has been a *loss*:

> No one can understand what it's like to carry this loss. In their minds, no baby meant no loss. (#212)
>
> To start with nobody got it. My grief was invisible. (#254)
>
> For the most part, the loss is unrecognized, there have been few people who have treated being childless as a loss and have acted sensitively. (#265)

Lack of understanding on the part of others can be compounded by one's own difficulties in comprehending and articulating the sense of loss, due to the absence of individual and shared interpretive resources. Initially, the experience might not be recognised as one of grief: "It took me a while to understand that what I was feeling was grief" (#214); "I never realised this was grief until I found others in a similar situation" (#254).

However, although ambiguity and lack of clarity might be especially salient in grief over childlessness, they are not specific to it. Similarly, vast swathes of a person's life could be affected by a significant bereavement, in ways that relate more or less closely to the death. Much of this remains invisible and intangible; it is not marked by shared practices, even if the death itself typically is. So, we might go so far as to say that unacknowledged losses are not only characteristic of occasional *disenfranchised* grief, but ubiquitous in bereavement and other challenging life circumstances (Rinofner-Kreidl, 2016). Hence, the sense of loss associated with bereavement can also be difficult to conceptualise and communicate, in ways that are not so evident if we take as our paradigm an emotional reaction to a death, abstracted from the context in which it occurs.

In the case of bereavement grief, we have noted that not all lost possibilities need be traceable to the event of the death. The term "anticipatory grief" is sometimes employed to refer to an experience of loss that occurs before someone actually dies (see, e.g., Sweeting and Gilhooly, 1990). We could think of this primarily in terms of a grief had in *expecting*, rather than responding to, a death. In other situations, though, anticipating something does not generally elicit the same kinds of emotions as its occurrence. We fear something before it has happened rather than afterwards, and we do not experience relief while anticipating it. However, it is often the case that certain possibilities are *already* lost before the death: we will no longer do that together; they cannot do this anymore. Similarly, those who are themselves dying may experience grief over what others will lose through their deaths (Varga and Gallagher, 2020). Thus, if we think of bereavement grief in terms of a wide-ranging experience of lost possibilities, it becomes clear how anticipatory grief can have a similar phenomenological structure to grief in response to a death. Conceiving of grief in this way also helps us to recognise the instability of its objects. The concrete objects of grief (and associated losses of possibilities) can include moving targets, as when experiencing grief over one's own or someone else's chronic progressive illness. As Kalanithi (2017, 16) observes, "the tricky part of illness is that, as you go through it, your values are constantly changing. You try to figure out what matters to you, and then you keep figuring it out".

Another feature shared by bereavement and other forms of loss is that both frequently affect how we relate to other people, and to the social world as a whole. Again, this is especially apparent in the case of involuntary childlessness. Being a parent involves being responsive to various shared norms, acting out established narratives, and participating in rituals, practices, and pastimes involving children. Lacking access to such possibilities can involve feeling "alone and different than others" (#208), in a world that continues to be structured around the anticipation and pursuit of parenthood. Survey respondents remarked, in particular, on finding "baby showers" emotionally difficult or unbearable. Others mentioned "daily reminders" from which there was "no escape", and how childlessness disrupts friendships with those who are parents. Analogously, in the case of bereavement, someone who has recently been widowed may find themselves faced with a world of couples, with friends who were always *our* friends, and with social activities that were always ways in which *we* interacted with others.

Construing bereavement grief as an instance of a broader phenomenological kind also reveals the complexities involved in individuating token grief experiences. It can be difficult to determine whether someone is experiencing a single wide-ranging loss or, alternatively, multiple losses. Suppose that someone loses their job, health, marriage, and then home. Are these to be regarded as objects of a single, unitary grief process that engages with the unravelling of one's life over a period of time, or as what Harvey and Miller (1998, 431) call a "pileup" of distinct losses? Importantly, the same question applies to bereavement, at least when we no longer consider it in abstraction from a larger situation. Worsened relationships with friends or family members, and also financial problems, are often occasioned by a bereavement (Eisma, de Lang, and Stroebe, 2022). Are these to be regarded as integral to a singular loss of possibilities?

It is arguable that a grief process should be construed as unitary to the extent that it engages with a unitary disturbance of life structure. A person's death can *imply* the unsustainability of an integrated network of projects, activities, patterns of thought, and interpersonal relationships (Ratcliffe, 2017a). Even so, we are doubtful that there is a sharp distinction between unitary and non-unitary losses, or between *experiences* of single and multiple losses. A grief process may involve experiencing losses in different ways at different times, sometimes as interrelated and sometimes not. Furthermore, a sense of lost possibilities could become more or less integrated over time.

5. Conclusion

We have set out a case for a broad conception of grief that encompasses bereavement grief and also a wide range of significant non-death losses.20 According to our account, grief involves experiencing – over a period of time – the loss of a network of more or less integrated possibilities that are integral to the structure of one's life. While accommodating a wide range of loss experiences, this conception still enables us to distinguish grief from the likes of disappointment and regret. Our account has implications for the understanding of bereavement grief. Although we do not seek to challenge more restrictive uses of the term "grief", which apply specifically to bereavement, we maintain that it can also be helpful to conceive of grief in more general

20 This conception of grief complements, and is corroborated by, empirical findings suggesting that clinically significant forms of grief (associated with labels such as "complicated grief" and "prolonged grief disorder") are common to both bereavement and other life situations. For instance, Shear et al. (2011) discuss "complicated grief" in the wake of Hurricane Katrina, in response to both bereavement and other losses such as losing a home or possessions. See Papa, Lancaster, and Kahler (2014) for further findings indicating that people who have suffered losses of various kinds can meet the criteria for prolonged grief disorder. Grief has also been reported in serious illness, as a response to various associated losses (Jacobsen et al., 2010).

terms. Doing so brings into view a phenomenological complexity that is overlooked when bereavement grief is construed as an emotional experience with a singular object and clear starting point, in abstraction from a larger situation. The acknowledgement of non-bereavement losses and their phenomenological structure also has important implications for whether and how we think about and respond to certain forms of human suffering. Without explicit, shared conceptions of these losses, they will shape human lives while evading established taxonomies, narratives, performances, rituals, and practices. Perhaps many of us are haunted in this way by inarticulate and unacknowledged feelings of loss, which amount to significant privations of the ability to sustain and develop who we are.

Acknowledgements. Research for this article was supported by the UK Arts and Humanities Research Council (Grant Ref. AH/T000066/1. We thank all of those who participated in our survey of grief experiences. We are also grateful to Jonathan Cole, Fredrik Svenaeus, and an audience at Södertörn University for very helpful discussions.

References

Attig, Thomas. 2011. *How We Grieve: Relearning the World*. Rev. ed. Oxford: Oxford University Press.

Boss, Pauline. 1999. *Ambiguous Loss: Learning to Live with Unresolved Grief*. Cambridge, MA: Harvard University Press.

Cholbi, Michael. 2022. *Grief: A Philosophical Guide*. Princeton: Princeton University Press.

———. 2019. "Regret, Resilience, and the Nature of Grief". *Journal of Moral Philosophy* 16 (4): 486-508.

D'Arms, Justin, and Daniel Jacobson. 2000. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions". *Philosophy and Phenomenological Research* 61 (1): 65-90.

Day, Jody. 2016. *Living the Life Unexpected*. London: Bluebird.

De Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, MA: MIT Press.

Di Paolo, Ezequiel A. 2020. "Comment: How Your Own Becoming Feels". *Emotion Review* 12 (4): 229-230.

———. 2021. "Enactive Becoming". *Phenomenology and the Cognitive Sciences* 20: 783-809.

Dodd, Celia. 2011. *The Empty Nest*. London: Piatkus.

Doka, Kenneth. J. 1999. "Disenfranchised Grief". *Bereavement Care* 18 (3): 37-39.

———, ed. 2002. *Disenfranchised Grief: New Directions, Challenges, and Strategies for Practice*. Champaign, Illinois: Research Press Publishers.

Eisma, Maarten C., Thomas A. de Lang, and Margaret S. Stroebe. 2022. "Restoration-Oriented Stressors of Bereavement". *Anxiety, Stress, & Coping*, *35*(3): 339-353.

Fuchs, Thomas. 2018. "Presence in Absence: The Ambiguous Phenomenology of Grief". *Phenomenology and the Cognitive Sciences* 17: 43-63.

Goldie, Peter. 2012. *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford: Oxford University Press.

Hadley, Robin A. 2021. "'No Longer Invincible': The Impact of Involuntary Childlessness on Older Men". *Physical Therapy Reviews* 26 (5): 328-343.

Hadley, Robin, and Terry Hanley. 2011. "Involuntarily Childless Men and the Desire for Fatherhood". *Journal of Reproductive and Infant Psychology* 29: 56-68.

Harbin, Ami. 2016. *Disorientation and Moral Life*. Oxford: Oxford University Press.

Harris, Darcy L. ed. 2020. *Non-Death Loss and Grief: Context and Clinical Implications*. New York: Routledge.

Harvey, John H., and Eric D. Miller. 1998. "Toward a Psychology of Loss". *Psychological Science* 9: 429-434.

Jacobsen, Juliet C., Baohui Zhang, Susan D. Block, Paul K. Maciejewski, and Holly G. Prigerson. 2010. "Distinguishing Symptoms of Grief and Depression in a Cohort of Advanced Cancer Patients". *Death Studies* 34: 257-273.

Kalanithi, Paul. 2017. *When Breath Becomes Air*. London: Vintage.

Klass, Dennis, Phyllis R. Silverman, and Steven L. Nickman, eds. 1996. *Continuing Bonds: New Understandings of Grief*. London: Routledge.

Korsgaard, Christine M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Marušić, Berislav. 2018. "Do Reasons Expire? An Essay on Grief". *Philosophers' Imprint* 18 (25): 1-21.

Mehmel, Constantin. 2021. "Grief, Disorientation, and Futurity". *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-021-09752-z

Moller, Dan. 2007. "Love and Death". *The Journal of Philosophy* 104: 301-316.

Papa, Anthony, Nicole G. Lancaster, and Julie Kahler. 2014. "Commonalities in Grief Responding Across Bereavement and Non-bereavement Losses". *Journal of Affective Disorders* 161: 136-143.

Ratcliffe, Matthew. 2008. *Feelings of Being: Phenomenology, Psychiatry and the Sense of Reality*. Oxford: Oxford University Press.

———. 2015. *Experiences of Depression: A Study in Phenomenology*. Oxford: Oxford University Press.

———. 2017a. "Grief and the Unity of Emotion". *Midwest Studies in Philosophy* 41: 154-174.

———. 2017b. *Real Hallucinations: Psychiatric Illness, Intentionality, and the Interpersonal World*. Cambridge, MA: MIT Press.

———. 2019. "Grief and Phantom Limbs: A Phenomenological Comparison". *New Yearbook for Phenomenology and Phenomenological Philosophy* XVII: 77-96.

———. 2020. "Towards a Phenomenology of Grief: Insights from Merleau-Ponty". European Journal of Philosophy 28: 657-669.

———. 2022. Grief Worlds: A Study of Emotional Experience. Cambridge, MA: MIT Press.

Ratcliffe, Matthew, and Eleanor A. Byrne. 2022. "The Interpersonal and Social Dimensions of Emotion Regulation in Grief". In Cultural, Existential and Phenomenological Dimensions of Grief Experience, edited by Allan Køster and Ester Holte Kofod, 84-98. London: Routledge.

Ratcliffe, Matthew, Louise Richardson, and Becky Millar. 2022. "On the Appropriateness of Grief to Its Object". Journal of the American Philosophical Association. https://doi.org/10/1017/apa.2021.55

Read, Rupert. 2018. "Can There Be a Logic of Grief? Why Wittgenstein and Merleau-Ponty Say 'Yes'". In *Wittgenstein and Phenomenology*, edited by Oskari Kuusela, Mihai Ometiţă, and Timur Uçan, 176-196. London: Routledge.

Rinofner-Kreidl, Sonja. 2016. "On Grief's Ambiguous Nature: Persisting vs. Transforming, Concealed vs. Expressed". *Quaestiones Disputatae* 7 (1): 178-207.

Sartre, Jean-Paul. (1943) 1989. *Being and Nothingness*. Translated by Hazel E. Barnes. London: Routledge.

Shear, Katherine M., Katie A. McLaughlin, Angela Ghesquiere, Michael J. Gruber, Nancy A. Sampson, and Ronald C. Kessler. 2011. "Complicated Grief Associated with Hurricane Katrina". *Depression and Anxiety* 28: 648-657.

Solomon, Robert C. 2004. *In Defense of Sentimentality*. New York: Oxford University Press.

Sweeting, Helen N., and Mary L. M. Gilhooly. 1990. "Anticipatory Grief: A Review". *Social Science and Medicine* 30 (10): 1073-80.

Teroni, Fabrice. 2007. "Emotions and Formal Objects". *Dialectica* 61: 395-415.

Varga, Somogy, and Shaun Gallagher. 2020. "Anticipatory-Vicarious Grief: The Anatomy of a Moral Emotion". *The Monist* 103 (2): 176-189.

Zahavi, Dan. 2007. "Self and Other: The Limits of Narrative Understanding". *Royal Institute of Philosophy Supplement* 60: 179-202.

Received: 15 August 2022
Accepted: 11 January 2023
Published: 15 June 2023

Framing the Role of Envy in Transitional Justice

Emanuela Ceva (University of Geneva) Emanuela.Ceva@unige.ch
Sara Protasi (University of Puget Sound) saraprotasi@gmail.com

Abstract

This article offers a conceptual framework for discussing the role of envy within processes of transitional justice. Transitional justice importantly includes the transformation of intergroup dynamics of interaction in the aftermath of societal conflicts and upheavals. Such transformation aims to realise "interactive" justice in transitional justice by reshaping belief and value systems, and by moulding emotional responses between the involved parties. A nuanced understanding of the emotions at play in intergroup antagonistic dynamics of interaction is thus essential to transitional justice. Among the many emotions that we could address in such scenarios, we target envy. Envy, in its various forms, features prominently in many societal conflicts and upheavals, and has, therefore, the potential to undermine or, conversely, support just intergroup interactions. However, the ambivalent role of this emotion has been scantly analysed in the philosophical literature on transitional justice. We make a start on filling this lacuna by developing a conceptual framework which is necessary to appreciate how envy and its varieties are epistemically and practically relevant to realising interactive justice in transitional justice processes.

Keywords: Envy, transitional justice, restorative justice, interactive justice, emulative envy, intergroup conflict.

0. Introduction

On August 10 1988, then US president Ronald Reagan signed the Civil Liberties Act. This act authorized the provision of a $20,000 lump-sum payment to all eligible persons[1] who had been adversely affected by an executive order signed by his predecessor, Franklin Delano Roosevelt, in 1942. Executive order 9066 authorised the forcible displacement of people of Japanese ancestry and their incarceration in isolated and secluded areas. From 1942 to 1945, 120,000 people, mostly American citizens, saw their human and civil rights violated: with only a few days' notice, they were forced to abandon their homes, leave most of their belongings behind, and to live in remote areas, in reconfigured buildings originally meant for livestock, without adequate sanitation or nutrition.

The last camp closed in 1946, but the executive act 9066 was only repealed in 1976, and it was not until 1988 that the US Congress would pass the Civil Liberties Act. The Act established monetary reparations to those

1 Only living survivors were considered eligible, and heirs of those who passed away before the act was signed were not eligible, so many victims and their families did not receive any reparations. We owe to Tamiko Nimura this and other clarifications about this case.

incarcerated, or their heirs, and a community education fund. In addition, the government issued a formal apology towards Japanese Americans, acknowledging the "grave injustice" that was done to them.[2]

From a philosophical point of view, the US government's actions in 1988 fall within the domain of *transitional justice*, standardly understood as "a legal and philosophical theory and a global practice aiming to redress wrongdoing, in order to vindicate victims, hold perpetrators accountable, and transform relationships" (Murphy 2021). The last element of this definition is crucial for us: transitional justice importantly includes the transformation of intergroup dynamics of interaction within countries or communities which have experienced societal conflicts and upheavals. Such transformation primarily aims to reshape belief and value systems, and to mould emotional responses between the involved parties. A nuanced understanding of the emotions at play in intergroup antagonistic dynamics of interaction is thus essential to transitional justice.

The intersection between the philosophical discussion of transitional justice and the philosophy of emotions has a certain tradition. A number of philosophical investigations into transitional justice processes have discussed the roles of emotions such as blame and resentment, on the one hand, and forgiveness and apology on the other (see, e.g., Blustein 2010; Elster 2003; Mihai 2016). But the emotional landscape of transitional justice is certainly more varied: guilt, hope, shame, regret, contempt, fear, hate, envy, and jealousy are all emotions that move perpetrators and victims of the type of societal conflicts and upheavals that are usually the target of this approach to justice.

We focus on one such important but unexplored emotion: envy. We define envy as someone's aversive, that is unpleasant or painful, response to a perceived disadvantage or inferiority vis-à-vis a similar other, with regard to a domain of self-importance, and which motivates the subject to either push themselves to the level of the envied, or pull the envied down to their level (cf. Protasi 2021; Lange and Protasi 2021). Envy is an inherently social emotion, whose concern is one's relative positioning (D'Arms and Jacobson 2006) and whose main evolutionary functions are to "facilitate successful social competition for access to resources that affect fitness" (Hill and Buss 2008, 63) and to regulate social hierarchies (Lange and Crusius 2022). Thus, envy need not be dysfunctional, imprudent, or immoral, even though it can be detrimental to a person's well-being and, more importantly for our discussion here, can wreak havoc in interpersonal contexts.

Let us now revisit the 1988 Civil Liberties Act, following the displacement and persecution of Japanese Americans during World War II. The context surrounding the Act cannot be fully explained without understanding the role played by a (racialised) form of envy towards the economic success of Japanese Americans.[3] These emotions might have faded in the intervening decades, and such fading might have favoured the reparations and apology issued in 1988, but they have not disappeared entirely, as the resurgence of anti-Asian hate in 2020-2021 has shown.[4]

2 As the conference report on this Act summarises, the Act declared that "(1) a grave injustice was done to citizens and permanent resident aliens of Japanese ancestry by the evacuation, relocation, and internment of civilians during World War II; (2) these actions were without security reasons and without any acts of espionage or sabotage documented by the Commission on Wartime Relocation and Internment of Civilians, and were motivated by racial prejudice, wartime hysteria, and a failure of political leadership; (3) the excluded individuals suffered enormous damages for which appropriate compensation has not been made; and (4) the Congress apologizes on behalf of the Nation" (U.S. Congress 1988). The same act also included reparations for, and an apology to, the Aleut people of the Aleutian Islands, who were also forcibly relocated during WWII to camps in Alaska, where many died of disease (Madden 1993).

3 This form of envy often pairs with racialised forms of jealousy and resentment toward those who were perceived as inherently inferior, and thus undeserving of their success, but we will not be able to address these other emotions here.

4 For an incomplete list of such hate crimes in the US, see Anti-Defamation League 2020. For a more global perspective, see Human Rights Watch 2020.

Analogous emotional phenomena, as we expand on in this article, can be found in past events as diverse (but also tragically similar) as the Jewish *Shoah* (Holocaust) and the Rwandan genocide. Furthermore, racialised envy plays a role in all-too-current socio-political upheavals that scholars of transitional justice are focusing on today, such as the search for racial justice in many Western countries, and, in particular, the issue of reparations for slavery owed to African Americans (see, e.g., Murphy 2021; for a different, future-oriented, approach to reparations, see Táiwò 2022).

While the idea that envy has an important role in antagonistic dynamics of intergroup interaction has some empirical support, no conceptual framework is currently available to understand how processes of transitional justice should study this emotion and its varieties. We devote our discussion in this article to developing such a framework, focusing on four components. In section one, we introduce the first component of the framework, which consists in the concept of transitional justice as transformation of the parties' dynamics of interaction. In section two, we then rely on the second component, the idea of "interactive justice", to explain the role of just interactions between parties in transitional justice processes. In the third section, we discuss a further component, consisting in the emotional underpinnings of the relational transformations implicated in, and required by, transitional justice processes. In section four, we introduce the final component, presenting envy as emerging out of an array of illustrative intergroup conflicts. In section five, we use our conceptual framework to illuminate how envy and its varieties may be epistemically and practically relevant to realize interactive justice in transitional justice processes.

1. Transitional justice as relational transformation

One possible way to see the Civil Liberties Act with which we started is as an instance of restorative justice (Takahashi 2013). Restorative justice is a process, within transitional justice, that prioritizes reparations for wrongdoing, and reparations in the form of apologies, compensation, and forgiveness (see, e.g., Philpott 2012).

There is, however, an emerging trend in the philosophical debate that shows the limits of reducing transitional justice to a restorative exercise (see Ceva and Murphy 2022; Murphy 2017). The core limit of such a reduction is that restorative justice implicitly presumes that acceptable common standards of interaction pre-exist the actual interaction between victims and perpetrators. Such standards may thus be taken as a guide for reparations. Yet the existence and content of such standards are disputable. Processes of transitional justice should recognize that the standards of interaction between parties need to be debated, and need *transformation*, not just *repair*.

Associated with this core limit, more specific concerns about restorative justice concern its (1) tendency to reduce parties to either victims or perpetrators, which risks exacerbating divisions and stigmatisation; and (2) targeting of compensations for wrongdoing rather than understanding the relationship dynamics from which the wrongdoing ensues (Ceva and Murphy 2022). Most importantly for our purposes, the emphasis that restorative justice places on forgiveness and reconciliation risks frustrating the expression of negative emotions. However, processes of transitional justice are densely populated by anger, fear, resentment, jealously, and, importantly for us, envy.

Before we can explain why studying envy is important to understanding antagonistic dynamics of intergroup interaction, as well as transforming them, some conceptual clarifications on the relational turn in transitional

justice are needed. Such clarifications will offer the first component of our conceptual framework for discussing envy in transitional justice processes.

The salience of a nuanced investigation into the role of negative emotions, in general, and envy, in particular, in transitional justice becomes apparent as soon as the transformation (not just the repair) of intergroup dynamics of interaction takes the stage. From this point of view, the subject matter of transitional justice is the quality of the relationships between subjects (usually belonging to different groups) implicated in a history of conflict, discrimination, and wrongdoing. The uniqueness of intergroup dynamics of interaction in transitional justice contexts justifies a relational focus. Such dynamics are tainted by structural inequality insofar as it concerns the terms of social interaction (Lu 2017), normalized oppression and human rights violations, and the ensuing uncertainty of the conditions of social cooperation and of the exercise of state authority (Murphy 2017). In such conditions, the quality of societal relationships becomes a distinctive concern of transitional justice in its own right.

The exact terms and conditions of the kind of transformation required to enhance the quality of intergroup dynamics of interaction may vary depending on the specific context where this process of transformation develops. Yet, by way of generalisation, we can posit that to study such processes means to discuss what may motivate parties who are emerging out of periods of conflict or wrongdoing to transform their antagonistic interaction into cooperative terms. Antagonistic interactions are characteristically tainted by the involved parties' mutual mistrust, which increases the emotional weight of dynamics of oppression, marginalization, and/or public stigmatisation (Deutsch 2000, 25-26). On the other hand, the marker of cooperative dynamics is openness to others, which means embracing a less obstructive emotional approach to interactions, which is capable of sustaining parties' mutual confidence and trust, despite the persistence of certain differences and disagreements. The philosophical study of transitional justice, from this perspective, consists in an analysis of the grounds on which parties may develop reasons for cooperation and how this change may be institutionally sustained.

To show how and why a relational transformation from antagonism to cooperation matters in the context of processes of transitional justice, consider, for example, processes of mediation. A mediator's role may certainly be assessed in light of the moral value of the results achieved by means of their action, for instance the quality of the compromise they helped the parties to negotiate. But mediation also has a transformative value (Bush and Folger 1994). Importantly, it requires fostering the parties' empowerment to seek solutions on their own, as well as reciprocal recognition. To this end, parties must overcome their mutual distrust and vent their claims in a protected setting, which may ipso facto change the way in which they look at their common past and their respective narratives.

Such a relational interpretation of transitional justice processes characterises the first component of the conceptual framework for our study. To understand and appreciate the affective dimension of transitional justice processes, the inherent properties of such processes should be the object of a specific investigation. Such an investigation should aim to analyse and assess the dynamics of interaction (that is, how the parties treat each other) that the process constitutes, or contributes to constituting, in society.

2. Just Interactions in Transitional Justice Processes

In the previous section, we have outlined a relational understanding of transitional justice. The analysis and assessment of transitional justice processes from this perspective focuses on the parties' just interactions. To wit, the core of this analysis and assessment is the just transformation of the parties' dynamics of interaction from antagonistic into cooperative terms.

To clarify, we refer to an interpretation of transitional justice as an instance of "interactive justice" (Ceva and Murphy 2022). Interactive justice focuses on the kind of treatment people are given and give to each other during their interactions in the context of institutional action (Ceva 2016). Through the lenses of interactive justice, the point of transitional justice is to set up processes capable of singling out unjust dynamics of interaction, analysing them, and understanding why they are wrong. In this way, parties may acknowledge and appreciate the terms of the relational wrongdoing implicated in their interactions. This is an essential step in changing antagonistic dynamics of interaction and, possibly, establishing cooperative ones instead.

The blueprint for such relational transformations may come from the experience of the many truth commissions established around the globe. Truth commissions are temporary bodies mandated to investigate the occurrence, causes, and impacts of specific human rights violations (Gonzalez and Varney 2013). One of the best-known processes of transitional justice of this kind was the South African Truth and Reconciliation Commission. This was established as part of the transition to democracy after decades of institutionalised segregation and racism under apartheid, from 1960 to 1994. Truth commissions are investigative, rather than punitive or reparatory. Their end results consist of a final report of the findings about human right abuses, with recommendations on how to avoid similar abuses in the future. Such findings substantially draw on the victims' own testimony, which contributes to rendering interactions during the hearings respectful of victims' voices and their agency. Testimonies are taken as prima facie valid, as contributions to establishing the truth, rather than as pieces of evidence for cross-examination, as in a criminal trial. This kind of interaction can show the parties involved how to relate to each other in more cooperative (respectful), and less antagonistic, terms.

Focusing on the relations between parties engaged in transitional justice processes was the first component of our conceptual framework; the second component consists in thinking about how to realise interactive justice in transitional justice processes by transforming those parties' relations in a just way, from antagonistic into cooperative terms. We can thus further specify our framework as thinking of transitional justice as interactive justice. Note that the transformation of the parties' dynamics of interaction occurs because of the new set of reasons for action that the parties may develop in virtue of their engagement in a transitional justice process. Because of their participation in the proceedings of truth commissions, for instance, voiceless oppressed groups may acquire the normative status of a victim of a wrongdoing and, consequently, gain the moral authority to stake certain claims, and the right to be heard. In the same process, the perpetrators of the wrongdoing acquire a duty to listen, and to develop a readiness to engage with the other side by reckoning with the other party's claims. The inclusion of such reasons in people's practical deliberative sets changes the way in which people look at and treat one another. What's more, participation in the process enables both parties to see beyond their status as bearers of conflicting claims. Transitional justice processes, therefore, realise interactive justice when they enable the parties to look at each other as parts of the same problem, and as equally important factors for imagining the solution to that problem.

One question that emerges out of this discussion concerns what may sustain the parties' commitment to acting on the new set of reasons for cooperation that participation in processes of transitional justice should be able to generate. Because antagonistic intergroup dynamics of interaction are characterized by a panoply of emotions, answering this question prompts a discussion of how transitional justice processes may purposefully mobilise this kind of emotional capital. Such a mobilisation is worth exploring in light of its potential to sustain or undercut the parties' commitment to acting on the cooperative set of reasons that accompanies their participation in the process.

3. Emotions in Transitional Justice Processes

The relational transformations that processes of transitional justice aim to realise are cognitively difficult and emotionally costly for the parties involved. To realise interactive justice, as discussed earlier, such transformations require a significant reshaping of the parties' own perspectives, of their understanding of their role in pivotal events in their own lives, and often in the life of a community or an entire country. Those who come to see themselves as perpetrators of significant wrongdoings may not only have to cope with newly-found guilt and shame, but also cope with loss of pride in what were previously considered honourable achievements. For example, this is often the fate of soldiers involved in mass atrocities or sexual exploitation and violence, but also of those who disown hateful value systems such as antisemitism or white supremacy.

The components of the conceptual framework we have outlined in the previous sections allow us to pinpoint the way in which the kind of transformation that transitional justice requires to realise interactive justice is not a by-product or a side effect of the process – it is the whole point. The emotional underpinnings of such a transformation may be an important resource for the process, as well as one of its main internal obstacles. Many emotional responses to such transformations and the context in which they occur may sustain or undermine the transformative process itself, and likewise the features of the ensuing dynamics of interaction between the parties. This consideration identifies the importance of exploring how transitional justice processes may purposefully mobilise a composite emotional capital in order to realize interactive justice. Such a mobilisation may take place either at the level of the individual or the collective, the latter perhaps even resulting in what William Reddy (2001) has called an "emotional regime", namely the set of norms governing emotional life which underpin stable political regimes. For instance, during a successful transitional justice process, all parties involved will come to see that rejoicing in an ethnic cleansing or enjoying a lynching is abhorrent, and that the appropriate emotional responses to such events are horror, moral disgust, grief, indignation, and so forth. In other words, transitional justice processes may bring "outlaw emotions" to the fore (Jaggar 1989) and counteract them. This feature of these processes may help to address and transform a complex emotional regime and, notably, make the involved parties recognise that certain emotional responses, but not others, are not simply acceptable but normative – that is, the type of emotions they *should* feel.

From the perspective of characterising our conceptual framework in terms of interactive justice, we can best appreciate how these momentous changes in emotional regimes are not only an outcome, but a precondition of, a successful transitional justice process. This is the case insofar as such changes can help sustain the parties' commitment to acting on the cooperative set of reasons that comes with their participation in the process. The capacity to appreciate the emotional underpinnings of the relational transformations that transitional justice processes realise is the third component of the conceptual framework of our study.

Before we introduce the fourth and final component of the framework, the conceptualisation of envy, combining the three components we have introduced thus far (the relational understanding of transitional justice, the idea of interactive justice, and the emotional underpinnings of relational transformations) prompts several groups of questions for the scholar of emotions who is interested in supporting the work of transitional justice. While we cannot develop each of them in full, we think it will be helpful to lay them out as a research agenda that our conceptual framework may sustain in the field.

First, and perhaps most obvious, is the question of the *fittingness* of the relevant emotions. We mentioned earlier how some previously unorthodox, but arguably fitting, emotional responses to lynching and mass murder can become normal and socially appropriate reactions. These are easy cases, presumably, from the perspective of any of the mainstream normative ethical theories. But there will be less easy cases, especially in the context of intergroup conflict. Some culturally-relative practices of a minority might elicit negative emotional responses in the majority, hindering reciprocal understanding. To give just two illustrations, consider the controversies surrounding the nomadic lifestyle of Roma people and other itinerant groups in Europe, or the practices of genital modification and mutilation performed in a variety of cultural contexts, with and without consent, on people of all ages, for an array of reasons, many of which tend to be hard or impossible to understand by people outside of the group that performs them (for a discussion of this, see, e.g., Galeotti 2007). To understand what the fitting emotional response to such practices is from outside of the group boundaries is a challenge in its own right.

The second question concerns the *epistemic* role of emotions, in addition to their *practical* action-guiding role. Emotions matter to transitional justice processes when it comes to both of these roles. First, investigating the emotions felt by the involved parties might help those parties to understand their dynamics of interaction better, shedding light on their motivations before and during their conflict. For example, recent philosophical research has analysed the risks and potential gains for the pursuit of justice of expressing anger in democratic public discourse (see, e.g., Cherry 2021; Lepoutre 2018; and Srinivasan 2018; *contra* Nussbaum 2015, 2016). In such contexts, those who listen to angry speeches can sympathise with the speakers, and gain a new insight concerning the claims of justice that the speaker's anger expresses. Besides this epistemic value, the public expression of anger may have practical import, too. Insofar as such an expression offers an enhanced insight into others' emotional landscapes, such insights may contribute to counteracting intergroup stereotypes, which would otherwise risk cornering the parties into opposed and irreconcilable positions. As discussed earlier, getting out of such corners is key to transforming antagonistic dynamics of interaction into cooperative ones, and thus to realising interactive justice in transitional justice.

This discussion, while cursory, shows the complexity of identifying the role played by the many morally relevant emotions in the processes of relational transformation required by the realisation of interactive justice in transitional justice. These emotions can have a productive or deleterious effect, and can themselves be affectively, morally, or epistemically positive or negative. What's more, some emotions that often populate interpersonal and intergroup interactions in these contexts are practically ambivalent. Depending on their specific connotations, they may therefore have the potential to sustain or hinder relational transformations.

4. Envy in Intergroup Conflicts

Broadly speaking, envy is a social emotion denoting someone's aversive response to a perceived disadvantage or inferiority vis-à-vis a similar other, with regard to a domain of self-importance (Protasi 2021; Lange and Protasi 2021). Envy has an ambivalent practical dimension insofar as, depending on its antecedent conditions, it motivates the subject either to push themselves to the level of the envied (constructive dimension) or to pull the envied down to their level (destructive dimension). Furthermore, again depending on antecedent conditions, envy's constructive or destructive tendencies can be paradigmatically sterile (leading to counterproductive behaviours in the envier), or fruitful (leading to the envier's successful emulation of the envied, or to the envied's loss of their advantage).

Envy's complex relational structure makes it a crucial candidate in refining the study of the emotional underpinnings of transitional justice processes. As we illustrate in this section, many conflictual histories of slavery, colonialism, racial segregation, and genocide suggest that envy is a constitutive component of the dynamics of interaction between the parties implicated in post-conflict or transitional scenarios.

Completing our framework with a conceptualisation of envy is thus important for two reasons. First, insofar as envy features prominently in the antagonistic intergroup dynamics of many conflicts, a framework for studying how those conflicts should be justly transformed must pay attention to the role of this emotion (inter alia) in fuelling the conflict. Second, because of envy's ambivalent practical dimension, the just transformation of the parties' antagonistic dynamics of interaction may usefully deploy strategies to defuse envy's destructive or counterproductive tendencies, and to strengthen its constructive and fruitful ones.

We outline the conceptualisation of envy through some examples, in accordance with the general characterisation of this emotion presented at the start of this section. These examples refer to some notorious intergroup antagonistic dynamics of interaction which characterise the Rwandan genocide, the internment of Japanese Americans, and anti-Black racism in the United States.[5] The diverse types and grounds of conflict that these examples illustrate are useful when it comes to appreciating the many facets of envy in intergroup dynamics of interaction, as well as the various ways in which this emotion may sustain or hinder the just transformation of such dynamics.

In the April of 1994, Rwandan President Juvenal Habyarimana's plane was shot down, an event for which the Hutu blamed the Tutsi minority. The day after, massacres of Tutsi and moderate Hutu began, and the nation descended into bloody chaos for four months. At least half a million Tutsi were dead by the end of it.[6] Many widows were raped and became HIV-positive, and about 400,000 children became orphans; many Hutus became refugees. An estimated one-third of the much smaller minority of Twa people was also killed.

To understand the events that escalated into civil war, take notice of some generally recognised factual elements concerning this case. German and Belgian colonial rule emphasised differences between Tutsi and Hutu, even though they actually spoke the same language and had similar cultural practices, regularly intermarried, and were physically very similar (so much so that, during the killings, many people were killed whose ethnicity

5 For the purposes of our discussion, we focus on conflicts involving (perceived) race and ethnic differences. However, note that envy may also feature in other kinds of conflicts, such as class and religion-based ones.

6 Estimates vary, depending on the source. Even the lowest estimate of 507,000 people represents two-thirds of the Tutsi population at the time (McDoom 2020).

was wrongly identified). Colonisers favoured the Tutsi by giving them administrative roles and higher places in society. In the 1930s, Belgian rulers officially divided the population into three ethnic groups (Hutu, Tutsi, and Twa), issuing compulsory identity cards. According to what we could call a colonial mythology, Tutsi were naturally superior, more intelligent, refined, and beautiful. After independence in 1962, however, the Hutu majority turned the tables, and subjected Tutsi to strict quotas in education and employment, thus giving Tutsi women an incentive to marry Hutu men so that their children "might escape such tight controls" (Hintjens 1999, 247). Although Tutsi had historically considered themselves superior to Hutu, they now found themselves discriminated against (albeit never segregated or ghettoised).

To look at such factual elements from the perspective of our conceptual framework is helpful in locating the deeply entrenched antagonistic dynamics of interaction between the two groups between 1962 and 1994, which run much deeper than their crude manifestations during the civil war in the 1980s and 1990s. Once such dynamics are in focus, we can also appreciate how any just process of transition away from the conflict must grapple with, and transform, such dynamics from antagonistic terms into cooperative ones, as interactive justice demands (see §2). What's more, our framework points out the importance of studying the emotional underpinnings of such antagonistic dynamics. Interestingly, even such critical voices as Helen M. Hintjens (1999), who resists the idea that the genocide itself was caused by intergroup antagonism, concedes that the conscious and deliberate political strategy that was mainly responsible, in her view, for the genocide, exploited and fomented such antagonism and its related emotional responses.

Among such responses we now want to highlight the role of envy. Indeed, the facts we have just reconstructed suggest that it is only natural that envy arose from both sides within the tense social and political circumstances leading to the civil war. While, to our knowledge, the role of envy in the Rwandan genocide has not yet been systematically analysed from a philosophical point of view, this emotion is mentioned in several discussions, both scholarly articles and anecdotal recollections of the events. For instance, envy is tracked as the basis of many reactions by the Hutu elites to the Tutsi's increasing economic and professional success (Hintjens 1999, 257; see also Clanton 2006, 438-439; Jean 2006; Nurhidayat 2022).[7]

Gender intergroup relations are also relevant in this context. For example, Hintjens writes about how Tutsi women were the object of sexual envy, because the wealthy ones were chosen as wives by Hutu political elite men after independence (see also Gallimore 2008, 20; Fielding 2014, 46).[8] Further support to the idea that envy did play an important role in shaping the Hutu-Tutsi dynamics of interaction via gender relations comes from Cristopher Taylor (1994, ch.4). Taylor discusses how Tutsi women were targets of both hate and sexual desire, often repressed, in the events leading up to and during the genocide. While Taylor does not use the word "envy" in his discussion, this emotion can be seen at work behind many of the gender norms and gender relations during that period. For instance, many powerful Hutu men had Tutsi wives or lovers, and a lot of anti-Tutsi propaganda clearly indicates that Hutu men had internalised the European stereotype that Tutsi women were more beautiful and intelligent. Thus, it is plausible that Hutu men envied Tutsi men, and that Hutu women envied Tutsi women (although Taylor focuses on the male part of this equation in his analysis). Both Hutu men and Hutu women likely perceived themselves as being in a disadvantageous position vis-à-vis

7 Franz Fanon, in *The Wretched of the Earth*, briefly explores the role of envy in analogous dynamics (Fanon 1963). Interestingly, in characterising the relationships between the colonised and the coloniser, Fanon speaks at length of envy, whereas resentment might be the more fitting emotion. We thank Samuel Elgin for this reference.

8 To reiterate, this was an inversion of previous hierarchies: "Tutsi aristocracy's erstwhile proprietorial use of Hutu women was not yet forgiven, and the sin was now seen as compounded by Batutsi women's supposed seduction of Bahutu elite men" (Hintjens 1999, 250).

Tutsi men and women, with regard to self-relevant domains (such as social status, personal attractiveness, and marital happiness). To bring envy into focus thus seems important to understanding central elements of the antagonistic dynamics of interaction during the conflict. It also seems a necessary component of any strategy for designing a just process of transition out of a conflict and towards the establishment of a cooperative form of interaction.

A somewhat similar dynamic of envy towards a social group that is both otherised and perceived as (economically) rising can be found in the events that led to the internment of Japanese Americans during World War II. Historical discussions of the Japanese internment suggest that the economic success of Japanese Americans was a source of envy, among other negative emotions (again, envy is not systematically discussed, but is often mentioned, when discussing the causes of the internment – see, e.g., Taylor 1999, 165; Burton et al. 2011, 27; for a similar point about Japanese Canadians, see Granatstein and Johnson 1988, 101).

While no systematic analysis is available in this case, , Protasi (2021) defends, for instance, the thesis that anti-Asian prejudice in the US has historically been imbued with envy. Reviews of anti-Chinese sentiment in the late 1800s reveal how initial praise of Chinese immigrants' work ethic slowly transformed into fear that these diligent labourers would steal jobs from Americans (see, e.g., Zia 2000). Immigrants from India were similarly initially welcomed and enjoyed the superior status of "Caucasians", until that privilege was revoked and they were relegated to the inferior condition of "non-white Caucasians" in 1923. A similar path of initial integration and subsequent marginalization and persecution can be seen in the aforementioned case of Japanese Americans, whose financial success was bound to cause envy, in addition to resentment. The myth of the "model minority" is rooted in this history of envious prejudice: Asians and Asian Americans are the target of ambivalent feelings which are centrally underpinned by envy, an envy that is primarily connected to seeing oneself as disadvantaged with regard to financial welfare and success, as well as academic and career achievements. Asians are not outright despised as other racial minorities are (who may be stereotyped as "lazy," "welfare recipients," "criminals" and the like), and yet, since this envy is racialised, and combined with the perception of the envied as "other" and foreign, it cannot easily be mobilised towards constructive and emulative goals, but rapidly and easily turns into maliciousness and hostility.

A recent empirical study by Akiba et al. (2021) shows that, at least in the recent wave of "Asiaphobia" (as the authors call it) following the COVID-19 pandemic, envy was a key motivator. Both Protasi and Akiba rely on the Stereotype Content Model developed by Susan Fiske, Peter Glick, and their collaborators (for an accessible review of the model, see Fiske 2011). The model predicts that Asians are perceived as unlikeable but competent, and are thus more likely to be the target of envy. To conceptualise envy and pay attention to its role may thus be helpful in clarifying the stakes in the dynamics of interaction between groups, even in more ordinary situations of social antagonism (and even in the absence of dramatic forms of unrest as in the Rwanda civil war), and, consequently, to transform such dynamics into more just (cooperative) terms.

Finally, to bring envy into focus may also help us to analyse and understand antagonistic dynamics of interaction involving the targets of a paradigmatically contemptuous kind of racism, such as African Americans. Even though Black people tend to be at the bottom of the hierarchical ideology of white supremacy, that does not mean that they cannot ever be envied, since the perception of a group's identity is always dynamic and context-dependent.

In many cultures, scorn for the alleged inferiority of an ethnicity coexists with envy and admiration of some allegedly superior traits. For instance, indigenous populations all over the world are stereotyped as inferior savages with lower intelligence, but at the same time as more creative and artistic, and with more primitive and less sophisticated – but more "exotic" – looks. In particular, Black Americans have been perceived in popular culture as more athletic, as more musical and better at dancing (and especially at certain types of music and dance, of course), and as hypersexual. This set of features has made them objects of envy. An example of this type of envy is portrayed in Spike Lee's 1989 film *Do The Right Thing*. In a famous scene, the Black protagonist Mookie remarks on how his racist white co-worker (and nemesis) Pino is a big fan of Magic Johnson, Prince, and Eddie Murphy. While Pino tries to explain the inconsistency away by suggesting that they are not "really" Black, Mookie suggests that Pino deep down wishes he were Black. This fictional scenario exemplifies a particularly ambivalent and complex instance of envious racial prejudice.

Taken together, the three examples that we have offered suggest that prejudice is always somewhat ambivalent, and that a dynamic of hate and desire could be traced in all of these scenarios, albeit manifested in different ways. To make reference to envy in the context of complex dynamics of intergroup interaction may be helpful to make sense of this ambivalence. This is because, as our working conceptualisation reveals, envy is inherently ambivalent, being destructive as well as constructive, counterproductive as well as fruitful. (We revisit this ambivalence in the context of transitional justice processes in the next section.).

To conclude this section: envy appears in a variety of intergroup conflicts. The cases mentioned in this section involve moral harm, even though there might be types of intergroup envy that cause only trivial harm (or no harm at all, if envy is benign, which we discuss below). In all of these cases, envy is felt by members of one group and directed towards members of another ethnic or racialised group. Sometimes this envy manifests itself within a personal relation that is strongly influenced by racial identity, as in Pino's envy toward Mookie. Envy underpins a number of different antagonistic intergroup dynamics of interaction, both as a collective or an individual affective state, with the two levels – collective and individual – being distinct but related. The integration of a nuanced theory of envy among the components of our conceptual framework may thus offer important insights about transitional justice processes, since the parties involved in those processes engage both in one-to-one and group-based interactions, and any such interactions may be the target of transformative initiatives.

5. Envy in Transitional Justice Processes

Through the examples in the previous section, we have suggested the importance of bringing envy into focus to analyse and understand many intergroup antagonistic dynamics of interaction. We now want to pinpoint how integrating a conceptualisation of envy into the analysis and assessment of transitional justice processes may help us to gain a better grasp of how such processes should be structured in order to transform the parties' antagonistic dynamics of interaction into cooperative terms, as interactive justice demands.

Because of envy's ambivalent nature, to look at processes of transitional justice through the prism of envy may be helpful in two ways, both of which our proposed framework contributes to highlighting. First, to study envy's destructive and counterproductive drives is significant in identifying an important (but currently under-theorised) element of the emotional underpinnings of the antagonistic dynamics that fuel many intergroup conflicts (whether more or less violent). Second, studying envy's constructive and fruitful

(e.g. emulative) effects may also contribute to understanding how the affective mechanisms typical of this dimension of envy can be mobilised to transform antagonistic dynamics of interaction into more cooperative (and therefore more just) ones.

Envy's ambivalence makes it a complicated emotion. Its complexity is heightened by its sneaky nature. Envy tends to hide and mask itself. Philosophers have remarked on this aspect of envy since antiquity (see, e.g., Plutarch 1959), and psychologists have provided empirical support for it (Miceli and Castelfranchi 2007). Since envy is unpleasant to feel, and necessarily involves the notion that one is comparatively inferior or disadvantaged, and furthermore because it is often a stigmatised and morally condemned emotion, enviers tend to not admit feeling envy, even to themselves. Sometimes, envy masquerades as resentment, a righteous emotion that signals an alleged injustice (while envy is amoral and need not concern any wrongdoing). Even jealousy enjoys a more reputable status, connected as it is to the notion of protecting a good that one already has.[9]

Envy's complexity suggests how unmasking this emotion, especially when potentially (or in actuality) malicious and destructive, is an important step for realising interactive justice in transitional justice. It is not possible to establish less antagonistic relations without addressing the green-eyed monster in the room, so to speak.

We must acknowledge that there is a degree of context-dependent variation in the concrete ways in which processes of transitional justice must address envy in its malicious forms, as well as mobilise the transformative potential of its benign varieties. Such variations by and large exceed the scope and boundaries of a philosophical discussion of transitional justice processes, as they depend on the empirical details of the various ways in which envy may manifest itself across intergroup dynamics of interaction. Notice also that even empirical works are scant in this department. Sociologists have overall neglected the study of envy. Even Gordon Clanton, the only contemporary sociologist who, to our knowledge, has written on the topic of envy in society, relies on either psychological evidence, dated sociological works (i.e., Schoeck 1966), or anecdotal evidence gathered by journalists (see, for instance, his discussion of envy in Russia in his 2006). Such empirical contextual considerations notwithstanding, we can offer a few suggestions about how our conceptual framework can illuminate our understanding of the role of envy in transitional justice processes.

First, even when envy comes in its malicious forms, when it underpins antagonistic dynamics of intergroup interaction escalating to civil war or genocide, this emotion retains its multifaceted connotation. Compare, for instance, the dynamics of envy in the context of the Rwandan civil war we discussed earlier to anti-Semitic envy. Jews have always been persecuted, marginalized, and literally ghettoised in Europe. Particular manifestations of envy towards an individual Jew, or groups of Jews, may have been fitting, and even non-malicious. A non-Jewish person might, for example, have envied their Jewish neighbour for accidental reasons, for example because of their greener lawn, without their envy being thereby antisemitic. However, the general sentiment of antisemitic envy is, by virtue of the vicious prejudice towards any Jew *qua* Jew that it embeds, always unfitting and very often malicious. Consider, now, the envy initially felt by Hutus during the early stages of the German colonisation of Rwanda, or even after independence in 1962, when antipathies and divisions were a lot more entrenched. In Rwanda, the divisions between ethnic groups were sowed by external colonial forces. The consequent envy was thereby based on very specific views, and a well-grounded sentiment of (purposefully created) intergroup competition. In this sense, Hutus' envy might have been fitting at least some of the time, in particular before the regime started enacting a programme of dehumanising propaganda against Tutsis (Hintjens 1999).

9 For a detailed philosophical discussion and review of the psychological literature on these topics, see Protasi 2021.

The framework we have developed in this article can bring the emotional underpinnings of the Hutus/Tutsis' dynamics of interaction into focus, and help to qualify the grounds of envy among such underpinnings. Our framework can thus offer important insights into the affective aspects that any process of transitional justice should address (and possibly counteract) in order to transform such dynamics in an interactively just manner. This exercise of transitional justice is important because it may enable the Rwandan population to look back at their history and acknowledge that they were pitted against one another, and that some of the negative emotions that arose from colonial practices were at least understandable, yet unfitting, because resentment, indignation, and anger at the colonial powers were the appropriate emotional responses, the emotions that should have been felt in response to the conditions they were facing. Those would have been the outlaw emotions that could have broken the emotional regime imposed by the colonial rule. Of course, it would be naïve to think that such acknowledgement is sufficient to heal a wound as deep as the one that affects post-1994 Rwanda, and it might well be too late now to overhaul a century-old rivalry. But thinking about these emotions might help future transitional processes to address similar, but less entrenched, conflicts.

What's more, having a framework for studying the role of envy in transitional justice processes is helpful in highlighting that, even when envy is malicious, it is a common emotion that need not lead to genocide and mass murder. What is more, this framework can also help us to see how to mobilise the most constructive aspects that, we have seen, characterise this ambivalent emotion. Among the constructive drives of envy, we should mention its capacity to motivate people and groups to redress injustice (Frye 2016), to help place focus on undeserved success (La Caze 2001), and to highlight what has value in a human life (Thomason 2015). Envy may also have functional roles in regulating social hierarchies and facilitating successful competition for limited resources. Indeed, recognising and understanding envy, even in its destructive and counterproductive varieties, has an important epistemic function in recognising and understanding certain social phenomena, and bringing to light their underlying motivations. Such a recognition and understanding is an important pre-condition for addressing and for transforming such phenomena.

We can now see how to discuss the many ways in which envy need not be malicious, or acted upon, is important to recognise this emotion and to aptly mobilise it to transform intergroup dynamics of interaction in the context of transitional justices processes. Destigmatising envy facilitates admitting this emotion, and thus contributes to having honest and forthcoming interactions between parties to transitional justice processes. This is an important desideratum for processes which aim to transform antagonistic dynamics of intergroup interaction in a cooperative sense. Of course, envy is not the only emotional underpinning of intergroup antagonism that needs addressing in order to enable the just transformation that transitional justice processes promise. However, to bring envy out into the open is an important contribution (which is still waiting to be realised) to such an endeavour.

We finally want to underscore a further sense in which it is important that the design and practice of transitional justice processes recognises the contribution that non-malicious varieties of envy may give to their transformative efforts. Recall the earlier discussion of racialised envy. When envy is imbued with racism, it tends to be dehumanising and destructive. But while racism, as a general attitude, is inherently unfitting and immoral, and thus requires complete eradication, envy in the context of racialised antagonistic interactions can be emulative and productive. Emulative envy focuses more on the good than on the envied, and perceives self-improvement as possible; thus, it motivates the envier to level up to the envied, that is, to push themselves to the envied's level in order to overcome their disadvantage (Protasi 2021). For instance, emulative envy towards the political achievements of other ethnic groups need not be destructive; it can, rather, motivate

one to fight for obtaining analogous political gains for one's own group – the flourishing of student unions modelled after Black student unions is an example of this pattern. To have a framework for studying such constructive aspects of the envy that underpins the dynamics of racialised intergroup interaction offers an important prop for the relational transformative efforts that characterise transitional justice processes as we presented them.

Furthermore, even when envy is not by itself emulative or otherwise benign, it can be a transitional emotion, paving the way to the transformation of antagonist intergroup dynamics of interaction into cooperative terms. For instance, during the 1992 Los Angeles riots, many stores owned by Korean Americans were ransacked and looted. At least in part, these acts were motivated by spiteful envy towards the economic success of these new immigrants. Asian Americans and African Americans have subsequently become political allies, and often see themselves as united in the same struggle against white supremacy and racism (see Protasi 2021, 146-148, for an expanded discussion of this example). In this way, being the target of similar envious reactions has, in some cases, favoured the development of a cooperative form of interaction between groups, as the realisation of interactive justice in transitional justice requires. To generalise, to the extent that cooperative relations are recognised as loci of interactive justice, as discussed earlier in the article, our framework allows us to appreciate the contributory role of (certain forms of) envy in realising a more just society via transitional justice processes.

Appreciating the contribution that the benign varieties of envy may give to the just transformation of intergroup dynamics of interaction should not leave us with too rosy a picture of such an emotion. We have already remarked on the ambivalent nature of envy. But even the benign varieties of this emotion, such as emulative envy, are not always supportive of interactive justice in transitional justice. For example, emulative envy is not always fitting. When, say, emulative envy is addressed towards a group whose superiority is grounded in an injustice (e.g., in a form of domination), the fitting affective response seems to be resentment. An illustration can be seen, for instance, in the type of affective reaction that the colonised may have towards the coloniser (see *contra* Fanon 1963). Some other times, emulative envy cannot arise, because self-improvement is not possible. In such cases, perhaps, less constructive varieties of envy may arise. To be able to recognise and address them as such (by bringing them to the fore during transitional justice processes) is nevertheless important for analysing the emotional underpinnings of antagonistic dynamics of intergroup interaction and, possibly, transforming them into more cooperative terms.

Nothing we have said in this article offers a conclusive argument for assigning a necessary and non-fungible role to envy in transitional justice processes. The aim throughout our discussion has been to develop a conceptual framework for the study of such a role. This framework, we have argued, is helpful to see why and how envy matters, and how it should be factored into the analytical study and normative assessment of the emotional underpinnings of intergroup dynamics of interaction, in contexts relevant to the realisation of interactive justice in transitional justice processes. Individuating envy's role in intergroup conflicts, differentiating between different varieties of this emotion, destigmatising it, and acknowledging the existence of constructive forms of envy can all help, epistemically and practically, to foster the relational transformations that transitional justice seeks.

To conclude, we submit that the most valuable contribution of the conceptual framework we have developed in this article is to offer the necessary conceptual toolkit for appreciating how envy is pervasive in intergroup conflicts and characteristic of many of the antagonistic dynamics of interaction which occur in those conflicts.

Therefore, admitting envy's existence among the emotional underpinnings of such dynamics is crucial for their enhanced understanding and transformation. Admitting envy allows the revelation of prior perceptions of inferiority and superiority (epistemic value), and movement towards establishing more balanced, symmetric, cooperative, and, therefore, interactively just relations (practical relevance). Insofar as establishing such relations is key to transitional justice processes, as we claim, to have a framework for investigating the role that envy plays in its varieties can illuminate the dynamics of interaction in intergroup conflicts and, thus, importantly contribute to realising transitional justice in society.[10]

10 The authors contributed equally to this paper, and are thankful to Alfred Archer and an anonymous reviewer for their helpful feedback and criticism.

References

Anti-Defamation League. 2020. "Reports of Anti-Asian Assaults, Harassment and Hate Crimes Rise as Coronavirus Spreads". *Anti-Defamation League* (blog). June 18, 2020. https://www.adl.org/resources/blog/reports-anti-asian-assaults-harassment-and-hate-crimes-rise-coronavirus-spreads.Akiba, Daisuke, Ana Sofia Velazquez Lopez, and Mari Hirano. 2021. "The Nature of Anti-Asian American Xenophobia during the Coronavirus Pandemic: A Preliminary Exploration into Envy as a Key Motivator of Hate". *Behavioral Sciences* 11 (11): 158.

Blustein, Jeffrey. 2010. "Forgiveness, Commemoration, and Restorative Justice: The Role of Moral Emotions", *Metaphilosophy* 41: 582–617.

Burton, Jeffery F., Mary M. Farrell, and Richard W. Lord. 2011. *Confinement and Ethnicity: An Overview of World War II Japanese American Relocation Sites*. Seattle: University of Washington Press.

Bush, Robert A. Baruch and Joseph P. Folger 1994. *The Promise of Mediation: Responding to Conflict through Empowerment and Recognition*. San Francisco: Jossey-Bass Publishers.

Ceva, Emanuela and Colleen Murphy. 2022. "Interactive Justice in Transitional Justice: A Dynamic Framework". *American Journal of Political Science* 66 (3): 762–774.

Ceva, Emanuela. 2016. *Interactive Justice*. New York: Routledge.

Cherry, Myisha. 2021. *The Case for Rage*. New York, NY: Oxford University Press.

Clanton, Gordon. 2006. "Envy and Jealousy". In *Handbook of the Sociology of the Emotions*, edited by Jan. E. Stets and Jonathan H. Turner, 410–442. New York, NY: Springer.

D'Arms, Justin, and Daniel Jacobson. 2006. "Anthropocentric constraints on human value". *Oxford Studies in Methaethics* 1: 99–126.

Deutsch, Morton. 2000. "Cooperation and Competition". In *The Handbook of Conflict Resolution. Theory and Practice*, edited by Morton Deutsch and Peter Coleman, 21–40. San Francisco: Jossey-Bass.

Elster, Jon. 2003. "Emotions and Transitional Justice". *Soundings: An Interdisciplinary Journal* 86: 17–40.

Fanon, Frantz. 1963. *The Wretched of the Earth*. New York, NY: Grove Press.

Fielding, Leila. 2014. *Female Genocidaires during the Rwandan Genocide: When Women Kill*. Hamburg: Anchor Academic Publishing.

Fiske, Susan. 2011. *Envy Up, Scorn Down: How Status Divides Us*. New York, NY: Russell Sage Foundation.

Frye, Harrison. 2016. "The Relation of Envy to Distributive Justice." *Social Theory and Practice* 42 (3): 501–524.

Galeotti, Anna Elisabetta. 2007. "Relativism, Universalism, and Applied Ethics: The Case of Female Circumcision". *Constellations* 14: 1, 91–111.

Gallimore, Rangira Béa. 2008. "Militarism, Ethnicity, and Sexual Violence in the Rwandan Genocide". *Feminist Africa* 10 (4): 21–34.

Gonzalez, Eduardo and Howard Varney. 2013. *Truth Seeking: Elements of Creating an Effective Truth Commission*. Brasilia: Amnesty Commission of the Ministry of Justice of Brazil; New York, NY: International Center for Transitional Justice.

Granatstein, Jack L. and Gregory A. Johnson. 1988. *The Evacuation of the Japanese Canadians, 1942: A Realist Critique of the Received Version*. Ottawa: Canadian Committee for the History of the Second World War.

Hill, Sarah, and David Buss. 2008. "The Evolutionary Psychology of Envy". In *Envy, Theory and Research*, edited by Richard. H. Smith, 60–70. New York, NY: Oxford University Press.

Hintjens, Helen M. 1999. "Explaining the 1994 Genocide in Rwanda". *The Journal of Modern African Studies* 37 (2): 241–286.

Human Rights Watch. 2020. "Covid-19 Fueling Anti-Asian Racism and Xenophobia Worldwide". *Human Rights Watch*. May 12, 2020. https://www.hrw.org/news/2020/05/12/covid-19-fueling-anti-asian-racism-and-xenophobia-worldwide.

Jaggar, Alison. 1989. "Love and Knowledge: Emotion in Feminist Epistemology," *Inquiry: An Interdisciplinary Journal of Philosophy*, 32 (2): 151–17.

Jean, Moise. 2006. "The Rwandan Genocide: The True Motivations for Mass Killings". *International Security*, 30 (1): 45–86.

La Caze, Marguerite. 2001. "Envy and Resentment." *Philosophical Explorations* 4 (1): 31–45.

Lange, Jens and Sara Protasi. 2021. "An Interdisciplinary Perspective on the Value of Envy". *Review of Philosophy and Psychology*.
Lange, Jens and Jan Crusius. 2022. "How Envy and Being Envied Shape Social Hierarchies". In *The Moral Psychology of Envy*, edited by Sara Protasi, 41–59. Washington, DC: Rowland and Littlefield.
Lepoutre, Maxime. 2018. "Rage Inside the Machine: Defending the Place of Anger in Democratic Speech". *Politics, Philosophy & Economics* 17: 398–426.
Lu, Catherine. 2017. *Justice and Reconciliation in World Politics*. New York: Cambridge University Press
Madden, Ryan H. 1993. "An Enforced Odyssey: The Relocation and Internment of Aleuts during World War II". PhD diss., University of New Hampshire.
McDoom, Omar Shahabudin. 2020. "Contested Counting: Toward a Rigorous Estimate of the Death Toll in the Rwandan Genocide". *Journal of Genocide Research* 22 (1): 83–93.
Miceli, Maria, and Cristiano Castelfranchi. 2007. "The Envious Mind". *Cognition and Emotion* 21 (3): 449–479.
Mihai, Mihaela. 2016. *Negative Emotions and Transitional Justice*. New York, NY: Columbia University Press.
Murphy, Colleen. 2017. *The Conceptual Foundations of Transitional Justice*. New York, NY: Cambridge University Press.
———. 2021. "How Nations Heal". *Boston Review*, January 21, 2021. https://bostonreview.net/articles/colleen-murphy-transitional-justice/.
Nurhidayat, Yusril. 2022. "The Rwandan Genocide: A Rwandan Perspective". *Social Science Works*. https://socialscienceworks.org/the-rwandan-genocide-a-rwandan-perspective/
Nussbaum, Martha. 2015. "Transitional Anger," *Journal of the American Philosophical Association* 1: 41–56.
———. 2016. *Anger and Forgiveness*. Oxford: Oxford University Press.
Philpott, Daniel. 2012. *Just and Unjust Peace: An Ethic of Political Reconciliation*. New York, NY: Oxford University Press.
Plutarch. (1st c. CE) 1959. *De Invidia et Odio*. vol. VII: Moralia. Loeb Classical Library.
Protasi, Sara. 2021. *The Philosophy of Envy*. Cambridge: Cambridge University Press.
Reddy, William. 2001. *The Navigation of Feeling*. Cambridge: Cambridge University Press.
Schoeck, Helmut. 1966. *Envy. A Theory of Social Behaviour*. Indianapolis: Liberty Fund, Inc.
Srinivasan, Amia. 2018. "The Aptness of Anger". *Journal of Political Philosophy* 26 (2): 123–144.
Takahashi, Rita. 2013. "Restorative Justice Almost 50 Years Later: Japanese American Redress for Exclusion, Restriction, and Incarceration". In *Restorative Justice Today: Practical Applications*, edited by Katherine S. van Wormer and Lorenn Walker, 225–232. SAGE Publications.
Táíwò, Olúfẹ́mi. 2022. *Reconsidering Reparations*. Oxford: Oxford University Press.
Taylor, Christopher C. 1999. *Sacrifice as Terror: The Rwandan Genocide of 1994*. Oxford: Routledge.
Taylor, Sandra. 1999. "The Internment of Americans of Japanese Ancestry". In *When Sorry Isn't Enough: The Controversy Over Apologies and Reparations for Human Injustice*, edited by Roy L. Brooks, 165-168. New York, NY: New York University Press.
Thomason, Krista K. 2015. "The Moral Value of Envy". *The Southern Journal of Philosophy* 53 (1): 36–53.
U.S. Congress. House. 1988. *Civil Liberties Act of 1987*. HR442, 100th Congr. Adopted 10 August, 1988. https://www.congress.gov/bill/100th-congress/house-bill/442.
Uvin, Peter. 1997. "Prejudice, Crisis, and Genocide in Rwanda". *African Studies Review*, 40 (2): 91–115.

Received: 15 August 2022
Accepted: 11 January 2023
Published: 15 June 2023

Affective Responses to Embodied Intelligence. The Test-Cases of Spot, Kaspar, and Zeno

Alessandra Fussi – alessandra.fussi@unipi.it

Abstract

The design, planning, and public presentation of service robots is gradually eroding the traditional distinction between social robots and service robots. This paper concentrates on one service robot (Spot) and two humanoid robots employed in the therapy of children diagnosed with ASD (Kaspar and Zeno). The paper argues that 1) Spot, which looks like a dog but is clearly identifiable as a machine, is less likely than Kaspar and Zeno to provoke the so-called uncanny valley effect. 2) The symbolic connotations associated with the robot's appearance are positive in the case of the two humanoid robots, while Spot elicits opposite responses depending on the context of use (e.g. trust in an industrial facility/fear and outrage when employed by the police in an underfunded neighbourhood). 3) Kaspar and Zeno play a positive role in a therapeutic setting, as affective mediators between neurotypical adults and children diagnosed with ASD, but the claim that they help children to recognise emotions is overstated. 4) Fears connected with weaponised robots built by other companies in imitation of Spot invite reflections on the mutual shaping between technology, society and politics.

Keywords: Social robots, service robots, ASD, emotion recognition, uncanny valley, situated affectivity

1. Introduction

It is well known that humans tend to project sociality when confronted with natural or mechanical entities that *appear* to be endowed with autonomous movement, either for the purpose of understanding their movement or in order to interact with them. For this to happen, a robot does not need actual social capabilities. It is enough that it seems to have them. Brazeal (2003) distinguishes three types of social robots, from those that are less sophisticated to those that are more advanced. To the first class belong machines that are "socially evocative." They have no real social capabilities, but they evoke the impression in users that they may have them. To the second class belong robots that have a "social interface." These machines know how to communicate by adopting relatively fluid language, matched with appropriate facial expressions and gestures. However, they are not receptive. Hence, they cannot properly interact with users. To the more sophisticated class belong machines that are "socially receptive." They can learn from exchanges with users, and are able to refine their communicative and interactive capabilities accordingly.

In this paper I argue that the traditional distinction between social robots and service robots is gradually being eroded in the design, planning, and public presentation of service robots. I do not concentrate on the actual interactive capabilities of robots (i.e., their strictly social skills). Rather, I discuss the affective responses elicited by two kinds of robots. My first example is a service robot named Spot, produced by Boston Dynamics, a company at the forefront of robot manufacturing, based in Waltham, Massachusetts. As I argue in sections 4 and 5, Spot is socially evocative in different important (and problematic) respects. My second example, which I discuss in section 3, concerns two robots with a social interface: Kaspar and Zeno. Kaspar is a child-sized humanoid robot developed by the University of Hertfordshire's Adaptive Systems Research Group; it was first produced in 2009, but has undergone several transformations since then. Zeno is a 43 cm tall humanoid robot developed by Hanson Robotics in 2007. They are both used for therapeutic purposes and research with children diagnosed with autism spectrum disorder (ASD). They show emotions, utter sounds and words, and move some body parts, but their features are simplified.

I will consider three factors that play a key role in the affective responses robots may elicit in the public. The first factor, which I will discuss in section 2, is the "uncanny valley effect," i.e., the cognitive and affective dissonance that may be provoked by machines that are very close to reaching lifelike appearance, but fail to replicate it with respect to behavior, fine motor skills, or perceptual cues. I will maintain that service robots like Spot are much less likely to elicit the uncanny valley effect than machines expressly built for social interaction, such as Kaspar and Zeno.

The second factor is the capacity to evoke affective responses thanks to the symbolic connotations that may be associated with the robot's appearance in different situations. By "robot's appearance" I mean the way it looks (for example: Zeno and Kaspar look like children), the kind of creature it evokes without necessarily being shaped like it (Spot evokes the appearance of a dog, but looks like a mechanical entity), and the way it behaves (does its movement appear autonomous? Does it inspire trust, or is it likely to elicit fear and anxiety? Can it be perceived as sociable?). As I will argue, Spot's association with a dog can give rise to opposing reactions. It can stimulate friendly feelings in a context such as an industrial facility, in which it takes the role of a docile and helpful companion, while it may elicit anxiety and fear in situations in which one could associate a dog's behavior with hostility and aggression.

The third factor is also contextual and situational. The production of social robots has heavily relied on research that aims to identify the emotions as episodic events that find expressions in certain facial configurations. This is the kind of research on which robots as therapeutic tools for individuals with ASD are based. However, philosophers of emotions and psychologists have long stressed that we cannot really attune to the emotion someone is expressing without taking into account his or her situation (Barrett et al. 2011; Griffiths and Scarantino 2009; Hess and Hareli 2016). Contextual information is essential to get the expression right (is someone about to cry because she is happy or because she is sad?), to understand why a certain emotion occurs (does she show anger because she wants to obtain compensation after being insulted, or because she missed her train?), and to grasp whether something in the environment is used as emotional scaffolding (is she listening to rhythmic music in order to get through her workout? Is she dimming the lights in her apartment to calm down after a rough day? Is she feeding her nostalgia by looking at certain childhood pictures? Is she feeling "on the wrong planet" as an autistic person?).[1]

1 There is an extensive literature on situated affectivity. See Stephan, Walter, and Wilutzky 2014 and Colombetti and Roberts 2014 for the application of the conceptual framework of situated cognition to situated affectivity; on emotional offloading see Colombetti and Krueger 2015; for music as a tool for emotional offloading see Krueger 2014. Scaffoldings for nostalgia are discussed in Massantini 2020. Krause-Jensen and Rodogno (2023) discuss the affective dynamics involved in the feeling of "being wrong" in people on the autism spectrum. Slaby (2016) argues that the user-resource model often employed in the analysis of situated affectivity does not sufficiently illuminate cases in which a normative domain constrains affective responses through attunement and habituation. He proposes "mind invasion" as a better conceptual tool, especially with respect to certain patterns in corporate workplaces.

As I argue, the thesis that therapeutic robots such as Kaspar and Zeno help autistic children recognise emotions is not supported by present research, besides being problematic at the theoretical level. The problem of understanding others' emotions cannot be reduced to guessing a robot's expressions, especially considering the cognitive style of individuals diagnosed with ASD (see Happé and Frith 2006). By analysing the results of some empirical studies, I infer that the benefits provided by these robots concern, at the perceptual level, the possibility to avoid overstimulation effects. With regard to the affective sphere broadly understood, benefits seem mainly to derive from the fact that these machines can be used as mediators between different cognitive styles (i.e., between that of the child with ASD and that of the neurotypical adult in a therapeutic setting). In this role, robots can alleviate social anxiety and facilitate shared attention, bodily awareness, or the exchange of roles in a game. I suggest that robots might also be able to bring benefits to the therapeutic setting by regulating the anxiety involved at both sides of the interaction (in this sense, they can be seen as affective scaffoldings). Precisely because their role as mediators would be lost if such machines were used outside of a therapeutic setting, for example as "companions" to leave alone with children diagnosed with ASD, in my concluding remarks I suggest that the potential disadvantages in this different situation should be carefully evaluated.

In my view, situational aspects are also key with respect to service robots. A problem that seems to be under the radar of public representatives and engineers is that affective reactions to embodied intelligence are not best studied in the abstract, or in a lab. One needs to consider the specific social environments in which service robots might be employed. I argue this point by contrasting the affective reactions elicited by Spot in different instances. In section 4 I analyse two videos posted at different times on the manufacturers' website. I argue that the first video intentionally blurs the distinction between Spot's actual capacities as a service robot and the social capacities it appears to have. The second video, by contrast, presents it as a powerful tool for gathering information, transporting materials, and moving into environments that would be dangerous or inaccessible to humans, but it focuses mainly on conveying a different message: Spot does not have a social interface and is incapable of autonomous movement. In section 5 I explain why, in my view, the second video can be read as a response to the events that led the New York Police Department to discontinue its contract with Boston Dynamics, following the reactions of fear and anger elicited by Spot when it was used by the police in the Bronx. In the remainder of the section and in the conclusion I highlight a further problem: machines like Spot can be weaponised. In this case, the context of interest is wider than the manufacturer of this specific machine in this particular time and place. My concluding point is that the question of whether service machines like Spot will bring more benefits than dangers must necessarily involve considerations about future scenarios in a globalised world. Hence the need to involve community members, users, philosophers of emotions, psychologists, and sociologists, so that problematic issues may become apparent before it is too late.

2. Anthropomorphic Projections and the Uncanny Valley *Effect*

As Hegel et al. (2009, 171) noted, certain conditions favour anthropomorphic projection: "Humans attribute human-like qualities to nonhuman agents depending on several parameters like appearance and social context – and consequently expect human-like actions."

There are limits, however, to anthropomorphic projections. Mori (1970) drew scholarly attention to a psychological phenomenon that hinders them. It became known, after the title of Mori's article, as the "uncanny valley" effect. One might expect a proportional relationship between technical refinement in the

production of robots that increasingly resemble human beings and the empathy and sense of trustworthiness they may generate in response. The relationship would be comparable to the one we experience when, pressing our car's accelerator, we expect it to increase its speed. However, Mori argued, the production of humanoid robots is better understood by analogy with mountain climbing: it is not true that the further you walk, the higher you climb, because, after reaching a peak, you are bound to descend into a valley. In the case of robots and prosthetics, the risk one runs after reaching the peak of extreme similarity to human shapes is to fall down into the uncanny valley. Minor differences indeed elicit feelings of repulsion. The consequence is that humanoid robots may end up hindering anthropomorphic projections precisely because they are too similar to the humans they are meant to resemble.

Mori identified two areas in which progressive similarity can reach critical peaks, with the transformation of a feeling of affinity into an eerie sensation:

1) Physical (static) resemblance. A wooden prosthetic hand is less similar to a human limb than a technologically advanced prosthetic hand in which nails and veins are carefully emphasised. A person shaking a hand of the latter type can be disturbed by the soft, boneless texture, or by the lack of warmth. All of a sudden the sense of affinity initially elicited by the way the hand looks is lost, and one finds it uncanny. This affective reaction is unlikely to happen in reaction to a well-crafted wooden prosthesis.
2) Functional mobility. A robotic arm unloading packages in a warehouse may be far less upsetting than a near-perfect humanoid robot that laughs while moving only certain facial muscles, speaks in a slowed-down manner, or does not move its eyes as one would expect.

The *uncanny valley* effect has been of great concern to those involved in digital animation. For example, the problem was found in the 2004 animated film *Polar Express*, directed by Robert Zemeckis. The movement and appearance of the characters is extremely realistic (much more so than in a typical Disney animated film of the same period). The accuracy and realism are achieved through the technique of "performance capture," in which an actor (in this case Tom Hanks) is filmed by a large number of computerised cameras as he acts in the scene. His movements are then transferred to the 3D models of the characters. The result in *Polar Express* is mostly excellent, but at times it causes a creepy effect. As is apparent in some of the frames, the protagonists are lifelike throughout, except for the detail in their eyes, which appear wide open, motionless, like those of the dead. In some cases, reviewers have complained about the lack of adequate facial muscle movement when characters speak. The realism of the figure is not matched by the micro-movements that characterise a living face. The motionless gaze, the reflection of light absent from the pupils, the eyelids not blinking at regular intervals, the rigid cheeks, are indeed disturbing.

Mori's basic suggestion is that if designers want their creations to elicit a feeling of affinity, trust, or even tenderness, as in the case of certain dolls or animated characters, they should avoid high degrees of human likeness, because in that case minimal elements of dissonance are sufficient to elicit the eerie effect. The original paradigm of the uncanny, according to Mori, is that of the corpse. In every way similar to the living body, it arouses deeply uncomfortable feelings once one notices its pallor or feels its icy temperature. It is no accident that the zombie and its variations are ubiquitous elements in horror films. And it is precisely the zombie effect to which some critical commentators trace certain images from *Polar Express*.

Damiano and Dumouchel (2018, 2), who echo Mori (1970), agree with him that a robot may look uncanny when a great resemblance to human traits is matched by uncoordinated or oddly mechanical movement. They point out, however, that the reverse is also true, namely, that if an object that does not physically resemble a human being behaves autonomously and in coordination with human movement, it provokes a social response, and people tend to interact with it as if it were human.

This asymmetry in anthropomorphic projection leads Damiano and Dumouchel (2018, 3) to contrast the affective response elicited by traditional dolls and that generated by robots. Traditional dolls are *ascribed* human traits thanks to a perceived physical resemblance. Robots, on the other hand, may look nothing like humans in physical appearance, but their movements may provoke the *inference* that they have human traits.

I find the authors' distinction between a projection based on static similarity and a projection based on motion similarity very interesting. It should be noted, however, that the second type of projection is not always traceable to inferential reasoning. Urquiza-Hass and Kortschal (2015), summarising the research of cognitive scientists and neuroscientists, observe that the cognitive system involved in identifying others as agents (and not just as external objects) is mostly not reflexive in nature. Rather, what is mostly at play is a (pre-reflexive, we might say) mechanism that involves areas of the brain deputed to social cognition. Citing the studies of Gallese and Rizzolatti, the authors recall that, being deputed to the imitation of movement, the mirror neuron system is also responsible for the synchronisation of movement and emotions in groups. Moreover, it serves to attribute intentionality not only within the human species, but also to animals, up to and including robots, especially when the observed actions belong to the repertoire of actions that agents can relate back to themselves. The recognisability of actions does not depend on the observed species, but simply on the similarity of their movement to one's own movement:

> The generality of this process indicates that the MN [mirror neuron] system may be less dependent on species-specific shape features than on general motor properties of subjects, animated such as in animals, or evoking the impression of animacy, such as in robots. Results of a functional magnetic resonance imaging (fMRI) study showed that when people observed motor actions of humans (talking, reading and biting), monkeys (lip smacking and biting) or dogs (barking and biting), the difference in activation of the motor and visual areas depended not on the species but on the actions shown (Buccino et al., 2004). Actions that are part of the observer's motor repertoire (talking, reading and biting) are processed by their motor system, including MN, while actions that are not in the observer's repertoire (lip smacking and barking) are processed based only on their visual properties. (Urquiza-Hass and Kortschal 2015, 169).

The thesis that actions such as eating, drinking, talking, and running are recognised by a system that does not reflect on the specific qualities of the object (e.g., similarity of shape), but automatically mirrors its movement, prompts the hypothesis that anthropomorphic identification is best achieved when reflexive modes of interpretation are not activated. Simply put, the less one reflects, the fewer obstacles there are to anthropomorphic projection. We can infer that anthropomorphic projection is favoured when, for a variety of reasons, conscious reflection is hindered, but at the same time there is also a sense of urgency to ascribe intentionality to certain movements.

Epley, Waytz, and Cacioppo (2007) argue that the urgency toward anthropomorphic mirroring is facilitated by three conditions: the first is cognitive, the second and third motivational. On the cognitive side, the authors

point out that one can normally rely on a wide range of knowledge and beliefs about humans in general, while little is known about nonhuman agents. It comes naturally to interpret problematic phenomena one encounters in different species, or even in mechanical objects, as if they belong to the human domain. The second and third conditions are *effectance* and *sociality*. *Effectance* consists in the need to act effectively in response to one's environment, and this also involves interpreting present stimuli and making correct predictions about the future. The attribution of human characteristics to nonhuman agents reduces the sense of uncertainty about the future behaviours of objects with which one interacts, and consequently increases one's confidence in the ability to respond appropriately.

The third condition, *sociality*, consists of the need to establish and maintain social relationships. It is not necessary to dwell at length on this type of need, which came dramatically to the fore during the COVID-19 pandemic. Since interaction and physical closeness are basic psychological needs, isolation breeds suffering, and a lack of social interaction provides a strong stimulus to anthropomorphism. The relationship of the main character in the film *Cast Away* with his personified friend offers an emblematic example of such a mechanism. Kelly (Tom Hanks), shipwrecked on a desert island, finds a Wilson Sporting Goods volleyball along with other debris from the crashed plane. At some point he draws a rudimentary face with his bloody handprints on the volleyball and starts treating it as his only companion. Talking to Wilson, expressing to it and sometimes at its expenses, the frustration of the castaway, saves Kelly from madness by allowing him to retain a connection to the world he has momentarily lost. Most importantly, anthropomorphic projection allows Kelly to cling to hope. He also persists in his many disappointing and frustrating attempts to go back to a life shared with others because his constant interaction with Wilson allows him to keep his social skills and desires alive (on the active power of hope and the difference between hoping and daydreaming, see Bovens 1999; Walker 2006; Meirav 2009; Miceli and Castelfranchi 2010).

Epley, Waytz, and Cacioppo (2007, 866) comment on the third motivation for anthropomorphism thus:

> In the absence of social connection to other humans. . . . People create human agents out of nonhumans through anthropomorphism to satisfy their motivation for social connection. This predicts that anthropomorphism will increase when people feel a lack of social connection to other humans and decrease when people feel a strong sense of social connection.

While it is true that anthropomorphic projection is most powerful when there is a suspension of the reflective process, when the burden of cognitive uncertainty is strong but one feels the urgency of certain answers, and when there is a deep need for companionship, it is evident that the manipulation of emotions by means of sensory stimulation has a key role to play in robotics.

Those who find themselves in the conditions described by Epley, Waytz, and Cacioppo (2007) are more ready than others to see, in certain movements similar to their own (such as walking, raising an object, following a moving object with their gaze, etc.), purposeful and affectively charged actions. Arguably, this is why the social robot industry often targets as potential users those who have not yet developed forms of complex rationality (children), those who suffer from isolation and affective deficiencies (the elderly), those who are cognitively impaired (dementia patients and some residents of nursing homes), those who feel isolated, and those who have problems in social interaction (for example, children diagnosed with ASD).

3. Learning Emotions from Kaspar and Zeno?

Given these premises, we might assume that the way social robots look (their static properties) has a less significant impact on users than the way they move (their dynamic properties). In general, however, producers of social robots tend to stimulate users' anthropomorphic projections by building machines that look humanlike, zoomorphic, or have a mixture of animal, human, and fantastic features.

Kaspar and Zeno are interesting from the point of view of their affective impact. They supposedly help children on the autistic spectrum to recognise emotions. Furthermore, they are used to enhance capacities for interactive and collaborative games, and to help modulate force used in interaction with others.

What does it mean that they help children on the autistic spectrum to recognise emotions? Behind this claim lies the theory originally developed by Ekman (see Ekman 1992; Ekman and Davidson 1994), according to which certain basic emotions are expressed by all of us without appreciable cultural variation thanks to certain micro-movements of the face. Basic emotions are schematised on Russell's scale along two axes: the horizontal axis of valence (pleasure/pain) and the vertical axis of activation (major or minor) (see Russell 1980; Russell and Barrett 1999). On this model (or *circumplex*) we find on the point of least arousal fatigue (or drowsiness), and on the point of greatest arousal surprise. At the negative valence extreme we see sadness; at the positive extreme, contentment. Along the quadrant we find some intermediate conditions: categorised with positive valence/low arousal are emotional states such as feeling calm, serene, relaxed; with strong arousal/positive valence joy, elation, excitement; with strong arousal/negative valence disgust, anger, fear; with low arousal/negative valence sadness and feeling depressed.

Kasper and Zeno essentially exhibit the emotions we find on the circumplex. One problem, however, stands out: their faces lack the complexity of micro-movements that characterises a living face, so that they are likely to elicit the "uncanny valley" effect. Videos posted on the web show therapy sessions in which children are asked by educators to recognise the emotions expressed by the robot's face. In a video recorded with Zeno, an educator asks a child to guess whether the robot is happy, sad, angry, or scared by comparing the machine's expressions with corresponding *emoticons* drawn on *flashcards* (ITV News 2017). The robot's facial movements are rudimentary and repetitive: the sad expression is always the same, as is the cheerful expression, the angry expression, etc. The premise is that the child, who might be disturbed by the frustrating complexity of the human face, can learn, in a safe environment and through repeated experience, to distinguish the robot's facial movements by giving them the exact name found on the flashcard. However, the limited range of facial movements may not only produce a perturbing effect (at least in neurotypical adults). It may make emotion recognition difficult. In one study dedicated to Zeno, Schadenberg et al. (2018) affirm:

> The emotions that can be modelled with Zeno's expressive face can be difficult to recognize, even by typically developing individuals. This can be partly attributed to the limited degrees of freedom of Zeno's expressive face, resulting in emotional facial expressions that may not be legible, but more importantly because facial expressions are inherently ambiguous when they are not embedded in a situational context. (Schadenberg et al., 2018).

The experiment described in the article consisted of adding to Zeno's facial expressions some vocal outbursts (such as "Wow!"; "No!"; "Ouch!"; etc.). As a result, the children's capacity to identify the emotions correctly increased significantly. One challenge that people with ASD face is integrating different sources of sensory

information (Kaliouby et al. 2006). In order to be helpful, the therapeutic setting needs to avoid on the one hand stimulation overload, and on the other hand extreme simplification. Sartorato et al. (2017) refer to several scientific articles documenting altered sensory perception in individuals with ASD (including overstimulation aversions). Problems seem to arise, more specifically, at the level of integration of multiple sensory modalities, while the processing of "simple" stimuli can be heightened (for example, the perception of pure tones). From a neurological point of view, multisensory stimuli are processed by allowing a "temporal binding window" (Sartorato et al. 2017, 3) so that stimuli that are not perfectly synchronous can still be ascribed to the same source. Apparently, in children with ASD the temporal binding window is larger compared to neurotypical children. This may explain the difficulty in accurately filtering stimuli from the environment, which might lead to sensory overload and anxiety, especially in crowded and noisy places (social gatherings, restaurants, theatres, etc.).

Affective cues in everyday life complement each other (e.g., vocal outbursts may accompany facial movements and changes in posture). Therapeutic machines cannot help in identifying emotions if they exhibit just one sensory cue (e.g., frowning), but they would be confusing if they perfectly matched the complexity of an emotional exchange in real life (e.g., frowning, shaking head, joking, saying something that is contradicted by a facial cue, using metaphors, moving eyes and hands, etc.).

The idea that a robot may teach emotion recognition raises two problems. The first concerns the cognitive style typical of children on the autism spectrum; the second concerns more generally the understanding of emotions in social relationships.

Regarding the first point, research has shown that a cognitive style described as "weak coherence account" is often encountered on the autism spectrum (Happé and Frith 2006). A person who can very accurately distinguish the parts of a whole may have difficulties when trying to identify the whole. For example, a situation may not be judged to be the same as a previous one if even the smallest detail changed, just as an object is not identified as that same object if a part of it is missing.[2] Different objects of the same type are only categorised as such with difficulty, because the perception of difference overrides that of similarity. For this particular cognitive style, each different example of a certain type (table) must be explicitly identified as such (this is a table, and also this, and this, and this, etc.), so as to generate a kind of large repository of images all referable to the same concept (I refer here to visual images, but the same can apply to sounds, tastes, etc.). Similarly, a typical situation can become a source of great stress if it changes with respect to one of its aspects (if you often go out to dine at a specific place, choosing a new restaurant may not be seen at all as fulfilling the promise "tonight we will dine at the restaurant").

Now, if a significant problem is recognising a whole when even minute parts have changed, it can be assumed that a child with this cognitive style will not be helped to recognise an emotion already seen when it reappears in the complexity of a living face, because in the movements of the new face many differences from those of the robot will be perceptible. The difficulties in this regard are numerous: a person's face changes very quickly when expressing emotions. Moreover, it certainly does not express only so-called basic emotions: it will be rare to find a person who expresses, in the same way as the robot, only contentment, only sadness, etc. Finally,

2 With respect to the integrity of objects, one may notice that it is not uncommon even among neurotypical individuals to be bothered when a favourite object undergoes minor accidents. For example, if one's favourite mug, still perfectly capable of holding liquids, falls to the ground and is a little bit chipped, some may react as if they could no longer recognise it as the object they were fond of using for tea.

people can express different emotions with the same expressions. One may certainly learn that when the robot raises its eyebrows in a certain way it expresses surprise, but someone who raises his or her eyebrows in daily life may communicate different things, just as someone who laughs may express joy, contempt, irony, surprise, nervousness, frustration, serenity.

Then there is a further problem related to the theory of emotions presupposed in the construction and therapeutic use of these robots. The majority of our affective responses do not happen in a social vacuum. While it is true that anger means different things depending on context, eliminating context also implies eliminating a fundamental element of understanding. Someone who is angry feels offended, but understanding someone means understanding *in what sense* they may feel offended. It involves sensing whether the other person feels belittled because their needs are not taken seriously, whether they think that values with which they identify are treated with contempt, whether their gender identity is not recognised, whether they find a certain behaviour unacceptable because they feel that their authority is threatened, etc. One certainly also makes contact with others by noticing the micro-movements of their faces, but one does not understand what they feel unless one encounters their world. Without a sensitivity exercised to understand the other's values, the imagination cannot explore the reasons that might be involved when someone displays expressions attributable to anger. To say, therefore, that the emotions manifested by the robot help in recognising similar emotions when they occur in reality seems to me to overstate the usefulness of the exercise.

I am not denying that interaction with the robot may bring benefits. Indeed, research shows that Kaspar can help with imitation, body awareness, and appropriate physical interaction (Costa et al. 2015), joint attention, and turn taking in a game (Huijnen 2016). Such benefits, however, could arguably derive from aspects other than the cognitive aid offered by recognising similar expressions in different faces. For example, it is possible to imagine that the anxiety of social interaction might be mitigated by the repetitiveness and reliability of the robot. If among the things that generate frustration and a sense of failure in the social world stands out its mutability and unpredictability, knowing that there is at least one playmate who manifests recognisable emotions and who does not get angry if these are not immediately understood or mirrored can provide peace of mind, and perhaps even help develop confidence and self-esteem. With stress removed, it may become less difficult to concentrate on certain tasks. Understood in this way, the robot functions as a transitional object (cf. Winnicott 1971): it does not make intrusive demands – for example, it does not demand to be looked in the eyes – and it offers stability, reliability, solidity, predictability. Furthermore, the proven benefits often presuppose a triadic social interaction. Costa et al. (2015, 269–70) describe the room arrangement with child and therapist sitting at the desk, both looking at Kaspar, which is placed on the desk at the apex of an ideal triangle:

> The arrangement of the actors involved in the session (robot, child and experimenter) had taken into consideration a cooperative position. In this arrangement of the room, two people work together on the same task, which provides an opportunity for eye contact and mirroring. The experimenter is able to move without the child feeling as if his territory has been invaded. Most importantly, this arrangement in a triangle allows the experimenter to encourage the child to engage in the interaction, without threatening his space and forcing eye contact.

Turn taking, joint attention, etc. are facilitated by the robot in a therapeutic setting, but humans set up, initiate, and maintain the interaction. It would be interesting to find out if the mediation provided by robots such as Kaspar and Zeno helps not just the children but also the therapists to regulate the anxiety involved in the interaction.

4. Dancing Robots, Busy Robots

In early 2021, a New Year's greeting video produced by Boston Dynamics appeared on several websites. The video, which was covered by major news outlets, depicts robots dancing to the sound of a 1962 song by the Contours called "Do You Love Me?" (2020).

The song's narrative is about a man rejected by his lover, apparently because he was clumsy and not fun to be around. He comes back, having become a skilled dancer, and provocatively asks, "Do you love me now that I can dance?"

Meanwhile, in the video, a lone humanoid robot begins to move a few timid steps to the rhythm of the music, and then gradually picks up the pace as robots of different kinds and shapes enter the scene one after another. In the end they all dance together. Their movements follow the music smoothly in a well-coordinated choreography. Arguably, the rejection evoked by the first words of the song is used in the video to hint at the sense of unease and fear inspired by the most primitive, awkward, and dangerous robots, which were the causes of even fatal workplace accidents. Bekey (2012, 20) cites a number of accidents caused by robotic arms used in industry, starting in 1979. The general public may have justified concerns regarding dangers posed by robots moving freely in an environment shared with humans. In response, the video seeks to inspire confidence by presenting robots that exhibit agility and coordination in a relatively small space.

The scene constitutes a perfect match to declarations one finds on the Boston Dynamics website. When asked if the company's robots use artificial intelligence, the answer takes as an example Spot, a steel quadruped robot, quite heavy (it weighs about 60 pounds), that is equipped with five cameras: two in front (the eyes), two on the sides, and one in the back. It resembles a dog and is a big hit for its mobility and ability to collect data in extreme physical situations:

> Out-of-the-box, Spot has an inherent sense of balance and perception that enables it to walk steadily on a wide variety of terrains. This form of AI that we call "athletic intelligence" allows Spot to walk, climb stairs, avoid obstacles, traverse difficult terrain, and autonomously follow preset routes with little or no input from users. (2023)

The expression "athletic intelligence" is philosophically significant, since it calls into play the claim that human thinking is not confined to intracranial activity (with the body conceived merely as an external tool that may or may not be employed). Rather, cognitive activity is situated, i.e., dependent on or even co-constituted by corporeality and the environment with which it interacts.[3]

Not only does the video reflect the image of athletic intelligence that is also made explicit elsewhere on the site, but it illuminates other significant aspects. Atlas, the humanoid robot, can dance far better than most of us, but, more importantly, it can do so with other robots (including Spot) by synchronising with their movements. As the title of the song suggests, robots not only ask for our love, but show that they deserve it, since they can build, together, a pluralistic, cheerful, open society capable of valuing different forms of situated corporeality and affectivity.

3 There is still no agreement on the different options regarding situated intelligence (co-constitution, co-dependence, and embeddedness). For an overview of the debate concerning embodied intelligence see Shapiro and Spaulding 2021. For a discussion of situated conceptions of intelligence in robot construction, see Duffy et al. 1999.

Two important points emerge from the video: on the one hand, artificial intelligence has reached unparalleled levels of development by working on movement, balance, and responsiveness to the environment. On the other, companies such as Boston Dynamics are developing service robots that are meant to appear sociable. In this regard, it is interesting to note that in May 2022 the company released a new video entitled "No Time to Dance" (2022). The primary purpose was to advertise new capabilities of Spot, which this time was pictured working in an industrial plant. Its cameras, as seen in the video, have been upgraded, and now use colour images instead of black and white as before. Spot is equipped with sensors and gyroscopes that allow it to take readings from gauges and send all the information it gathers back to the computer. It can be guided with a tablet or programmed to perform a mission on its own. Significant upgrades include a new tablet, more sensitive and rugged than the previous one, and better batteries, with faster charging times. Payloads may add additional sensors and cameras, as well as the possibility of 5G connectivity.

The video is interesting beyond the information it gives about Spot's new capabilities, since it refers back to the previous narrative and slightly modifies it. Spot has so much work to do that it cannot waste its time dancing – or so we learn from two employees who guide it remotely with their tablets, while a technician, in a different room, analyses the data collected by the robot. The two employees are amused by a man they observe from above, a worker who (like many of us, one may assume) was mesmerised by the previous Boston Dynamic video and follows the robot around the industrial facilities trying to persuade it to dance with him. Spot is plainly incapable of interacting with the man, who refuses to accept that what he has before him is neither a real dog nor a social robot: he pats it on the back, does a few dance steps hoping it will imitate him, even offers it some bolts to eat (what food can a robot dog like if not something made of steel?). Eventually Spot starts dancing with him, but this time we are made aware of what is really going on. Without the worker realising it, the robot is remotely guided by the two employees, who exchange amused and patronising remarks about the worker, while he is left under the illusion of dancing with his new friend. Viewers who may have formed opinions similar to those ascribed to the naive worker are informed by this new video that Spot's autonomy is in fact very limited: it can be programmed to repeatedly follow a certain path or be guided by people using a tablet. It is not a machine meant for social interactions. Furthermore, it is plainly incapable of doing anything on its own initiative.

Before asking ourselves why the company might have chosen to send this new message to the public, let us remind ourselves of the difference between the capacity to socially interact with human beings and the capacity to engage with other robots. A "society of robots" was, after all, evoked in the "Do You Love Me?" video.[4]

The idea of a society of robots is discussed by Bicchi and Tamburrini (2015, 239–40), who interpreted the phenomenon in light of a sociology of subculture(s). They proposed two examples: the first concerns a group of robots employed in museums as night and day guards, the second is about a group of robot butlers working in a shopping mall. The robots must be able to communicate with each other while at the same time avoiding collisions with objects, moving carts, and the crowds of visitors. It is sufficient for such robots to behave as social groups, without being socially responsive towards human visitors. They do not communicate other than within their group; they have their own language and their own codes of behaviour.

4 On the difference between *societal robotics* and *social robotics*, Duffy et al. (1999) maintain that "the former represents the integration of robotic entities into the human environment or society, while the latter deals specifically with the social empowerment of robots permitting opportunistic goal solution with fellow agents"; Fong et al. (2003) describe the evolution from societies of robots inspired by insect societies (self-organised, homogeneous groups formed by anonymous individuals) to societies formed by heterogeneous groups (robots and humans) in which individuals can recognise each other, communicate, socially interact, and learn from each other. This model requires the construction of "individual social" robots (see Duffy 2004). For further reflections on social robots, see Hegel et al. 2009; Naneva et al. 2020.

From a certain point of view, the image of a subculture may cause apprehension: a subculture may seem a short step away from becoming a counterculture that does not share the dominant values, opposes them, and may seek to subvert them. On the other hand, precisely the unease that this image evokes reflects the current affective ambivalence toward embodied artificial intelligence. If it is becoming more and more usual to see individual robots used in industry or in the service of individuals (think of the millions of Roombas vacuuming up dust and moving around in our homes), the idea of a robot society may be frightening. One can imagine, then, that a first step toward generating acceptance in the public would be for manufacturers to envision groups of robots that refrain from explicit interactions with humans and make the communicative relationships they have with each other as inconspicuous as possible.

Such robots must have an appealing appearance, fluid movements, the ability to collect and transport data and objects, and, above all, the ability to avoid humans. They need to be able to calculate their trajectories in order to change direction whenever there is a risk of getting too close or even colliding with humans or cutting them off. Significantly, one of the applications Spot works with causes it to abort its mission as soon as it gets (too) close to a person. Still at the beginning of the path to their integration, robots must move in the background of human life. They need to look like dumb, docile, harmless servants.

5. Friendly Dogs, Scary Dogs

The interactive capacities of social robots seem to be on the whole still underdeveloped (Henschel et al. 2021), while in some cases service robots have very strong potential for eliciting anthropomorphic projections. Spot is a service robot endowed with characteristics that, as we have seen so far, tend to elicit positive reactions. It does not produce the uncanny valley effect because it does not aim to be mistaken for a living dog: it is clunky, rather noisy, it is made of steel, and does not hide its mechanical structure. For all these reasons it does not give rise to dissonant effects at the level of perception.

However, we noted that the "No Time to Dance" video issued by Boston Dynamics addresses precisely some unrealistic expectations Spot may provoke. This machine, we learn from the video, is incapable of autonomous movement. In contrast with a real dog, it cannot act without human control. It can either be moved from a distance by a tablet, or it can be programmed to follow a certain path and to perform a definite set of actions. Furthermore, Spot is clearly not a machine designed to interact with humans: it is not a social robot.

Why did Boston Dynamics feel the need to lower the public's expectations concerning this incredibly efficient machine?

The reason may lie in the polemics that followed the employment of Spot by a few police departments in the United States. In particular, when the New York police department bought a few exemplars (renamed Digidog) and started using them in areas of the city that usually fall under the radar of public funding, the backlash produced by their employment ultimately led to their dismissal. Public outrage began when, in February 2021, the NYC police employed Digidog during an operation in the Bronx, as reported by the New York Times (Cramer and Hauser 2021). Apparently, two men armed with guns were holding two other men hostage in an apartment. When policemen arrived at the scene, they employed Digidog to explore the building. The inhabitants reacted with outrage and fear when they saw the mechanical dog let loose in their neighbourhood.

Political reactions immediately followed. Representative Ocasio-Cortez, for example, called into question the facting around $70,000) were given precedence over meeting much more urgent needs in the Bronx. On Twitter, she said:

> Please ask yourself: When was the last time you saw next-generation, world-class technology for education, health care, housing, etc. consistently prioritized for underserved communities like this? (2021)

In a New York Times article published on April 14, 2021, journalist Mihir Zaveri reported on a new incident involving the use of Digidog in a public housing building in Manhattan. Apparently, the robotic dog did not play any active role in the apprehension of a man who, armed with a gun, was keeping a woman and a baby hostage in one of the apartments. However, people who saw the robot in the lobby reacted with fear and outrage at the use of public money for this kind of device (Zaveri 2021).

The political debate that ensued focused on the lack of transparency concerning the police department's surveillance tools, the likely breach of privacy, the possibility that tools of this sort could reinforce police bias against underprivileged groups, and the likelihood that such machines could be weaponised and developed to initiate autonomous action.

One of the most frequent associations evoked by commentators was with the robotic dogs appearing in a widely popular 2017 episode of Black Mirror, "Metalhead." The episode portrayed a dystopian society in which robotic dogs lacking any capacity for empathy were chasing and killing humans who were trying to recover a teddy bear for a child who desperately missed it. The resemblance between Digidog and the robot dogs appearing in "Metalhead" is unmistakable. Indeed, the series creator, Charlie Brooker, affirmed in an interview (Hibberd 2017) that he took inspiration from a Boston Dynamics video (posted on YouTube in 2015) portraying a robot dog being kicked.[5] While the video was meant to showcase Spot's capacity to maintain balance in extreme situations, several viewers posted comments showing empathic distress at the machine being abused by humans. "Metalhead" takes the opposite affective perspective: it shows how threatening a machine like that can be.

In response to the polemics following the employment of Digidog by the New York Police Department, representatives of Boston Dynamics clarified that their devices are mostly used in industrial facilities and in commercial places where they can replace humans in dangerous settings. In February 2021, however, an art collective mounted a paintball gun on a Boston Dynamics dog, allowed participants to take control of it remotely, and let it shoot up a gallery (the event was significantly named "Spot's Rampage"). In response, Boston Dynamics circulated a statement on Twitter condemning the portrayal of its products in ways that promote "violence harm and intimidation" and reminding the public that all buyers of the company's products must comply with Terms of Agreement that prevent their robots from being used "to harm or intimidate people or animals." Violations of the Terms of Sale would void the warranty and make it impossible to update, repair, or replace the products which had been misused.[6]

5 This video (accessible at https://www.youtube.com/watch?v=aR5Z6AoMh6U), titled "Watch robot dog 'Spot' run, walk...and get kicked" (runtime 1:11), was posted by On Demand News on 11 February, 2015.

6 Boston Dynamics (@BostonDynamics), Twitter, February 20, 2021, https://twitter.com/BostonDynamics/status/1362921918781943816.

In light of events that led the New York Police Department to terminate its contract with Boston Dynamics in April 2021, we may understand why the company's most recent videos aim to lower people's expectations about Spot, underlie its usefulness mostly in industrial facilities, and stress that it lacks the capacity for autonomous movement.

The company's terms of use follow ethical principles posted on the website (n.d.). One of the principles states the following:

> We will not authorize nor partner with those who wish to use our robots as weapons or autonomous targeting systems. If our products are being used for harm, we will take appropriate measures to mitigate that misuse.

Since science does not happen in a vacuum, however, other companies started building quadruped robots with Spot's capabilities. Some did indeed weaponise their products. A company called Ghost Robotics (based in Philadelphia) presented a robot dog with a gun mounted on it at the Association of the United States Army's 2021 annual conference in Washington DC. Similar weaponised dogs were produced in China and in Russia. In an open letter to the robotics industry, Boston Dynamics pledges neither to produce nor to support weaponised mobile robots, and issues a call "on every organization, developer, researcher, and user in the robotics community to make similar pledges not to build, authorize, support, or enable the attachment of weaponry to such robots." One of the reasons adduced in the letter is that weaponised robots hinder public trust in robotic machines and obscure the fact that they can bring great benefits to society if used properly. I have no reason to question the sincerity of Boston Dynamics' pledge. However, I doubt that it will seriously influence the process of weaponisation of robot dogs.

6. Conclusions

Anthropomorphic projections towards embodied intelligence can be sympathetic, but they can also be very hostile. Industries like Boston Dynamics focus on the technological capabilities of their machines and issue products that are both carefully designed and skilfully promoted. However, something clearly went wrong in the case of Spot.

We saw three problems: the first arose from the associations that, via "Metalhead," made Spot look scary. The second was due to the lack of a fruitful exchange between the police, the company, and political and social representatives concerning public policies. Dogs can be good to friends and nasty to enemies, but of course they can be trained to identify as enemies anybody one may choose as a target of aggression. This much was known since Plato's *Republic*, in which the warrior class was associated with dogs, both in its positive qualities and in its more dangerous side (education of the warrior class was meant to limit the danger of an armed class that could attack its own citizens instead of defending them; see Fussi 2022). The third problem concerns the possibility that machines like Spot may be weaponised.

While the uncanny valley effect plays a role at the level of individual perception, the second and third problems are clearly situational and contextual. A service dog like Spot can elicit sympathy in an industrial setting, fear when it is enrolled by the police, outrage when it is used as a model for new instruments of destruction and social control.

The first conclusion we can draw is that industries developing embodied intelligence should consider carefully the symbolic connotations that may be associated with the designs chosen for their robots. The second is that the introduction of embodied intelligence in human environments should involve not just scientists, legal experts, and ethicists during the production process, but also social scientists, political representatives, as well as psychologists and philosophers of emotions.

The third conclusion is that it is necessary to think long and hard before arriving at the production process, about the possibility of future uses and misuses of technologies. It cannot be beyond imagination that a robot dog, guided from a distance, endowed with cameras and refined sensors, capable of carrying weight and moving in all kinds of terrain and temperatures, may become an instrument of death. Endowed with software for facial recognition and powerful cameras, a machine like that can be used by organisations both interested in surveillance and indifferent to privacy laws or civil rights. The context in which scientists and companies make decisions cannot be confined to their labs, to their good intentions, or to their own wish to instil public confidence in the machines they produce here and now.

At first sight, the examples we discussed in this paper may appear very different. Spot is a service robot, while Kaspar and Zeno are therapeutic robots. Spot does not have a social interface, while Kaspar and Zeno do. However, Spot is less likely to elicit the uncanny valley effect than the two humanoid robots that are used with children diagnosed with ASD. From one point of view, Spot is also much more likely to invite social projections (as we have seen both by analysing the videos provided by Boston Dynamics and by following the events that led to its dismissal from the New York Police Department).

An overly simplified interpretation of what it means to understand emotions can lead us to overstate the role played by therapeutic robots like Kaspar and Zeno. Their main function may be found, not in enhancing the capacity to recognise occurring emotions, but in facilitating affective interactions with others in a triadic social setting. As I suggested, in such settings it is likely that not only the children involved, but also their therapists, may regulate the anxiety involved. If this intuition proved accurate, we could say that therapeutic robots are affective scaffoldings and function as mediators between different cognitive and affective styles.

Would it be as advantageous for a child who has difficulties with social interaction to be able to spend a lot of time alone with Kaspar? Could this not reinforce isolation, repetitiveness, aversion to the complexity of social involvement with other human beings? This is another problem that needs urgently addressing, given the announcement recently posted by the University of Hertfordshire:

> Following positive results from field trials in schools and homes – and ahead of an upcoming NHS trial – the Kaspar research team is looking to redesign and reengineer its Kaspar prototype into an advanced, robust and commercially viable robot that can be made available to any child around the world that needs it. (n.d.)

Ultimately, I agree with Šabanović (2010) when she points out that it is too often the case that scientists consider how societies can adjust to technological improvements instead of considering the mutual shaping between technology and society. Producers of robots should involve citizens and their political representatives in the early stages of design, rather than expect them to passively adapt to their products once they are on the market. Cultural and political connotations should be taken into account, and the different contexts in which robots will be used ought to play a significant role in a bottom-up approach to robot production. Šabanović

(2010) mainly addresses the issue of social robots. However, as I hope becomes clear from my examples, the affective responses elicited by sophisticated service robots like Spot suggest that this sort of strategy ought to be extended to embodied intelligence in general, especially when public funding and public policies are at stake.

Emotions such as fear, hope, and anxiety play a role in the relatively small social settings we examined when we considered Kaspar and Zeno, as well as in the differentiated public settings we evoked in the case of Spot (industrial facilities, police forces, city neighbourhoods, military conventions, art galleries). Such emotions concern the future. Inspired by them, we urgently need rational, imaginative, collaborative thinking on future ecological settings inhabited by humans and robots.

References

Barrett, L. F., B. Mesquita, and M. Gendron. 2011. "Context in Emotion Perception." *Current Directions in Psychological Science* 20 (5): 286–90.

Bekey, G. A. 2012. "Current Trends in Robotics: Technology and Ethics." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by P. Lin, K. Abney, and G. A. Bekey, 17–34. MIT Press.

Bicchi, A. and G. Tamburrini. 2015. "Social Robotics and Societies of Robots." *The Information Society: An International Journal* 31 (3): 237–43.

Boston Dynamics. n.d. "About." Accessed May 20, 2023. https://www.bostondynamics.com/about#Q3.

———. n.d. "Boston Dynamics Ethical Principles." Accessed May 20, 2023. https://www.bostondynamics.com/ethics.

———. 2020. "Do You Love Me?" 29 December, 2020. Video, 2:54. https://www.youtube.com/watch?v=fn3KWM1kuAw.

———. 2022. "No Time to Dance." May 3, 2022. Video, 2:57. https://youtu.be/6U-bI3On1Ww.

Bovens, L. 1999. "The Value of Hope." *Philosophy and Phenomenological Research* 59, 667–81.

Brazeal, C. 2003. "Emotion and Sociable Humanoid Robots." *International Journal of Human Computer Interaction 59: 119–55.*

Buccino, G., F. Lui, N. Canessa, I. Patteri, G. Lagravinese, F. Benuzzi, C. A. Porro, and G. Rizzolatti. 2004. "Neural Circuits Involved in the Recognition of Actions Performed by Non-Cospecifics: An FMRI Study." *Journal of Cognitive Neuroscience* 16 (1): 114–126.

Colombetti, G. and J. Krueger. 2015. "Scaffoldings of the Affective Mind." *Philosophical Psychology* 28 (8): 1157–76.

Colombetti, G. and T. Roberts. 2014. "Extending the Extended Mind: The Case for Extended Affectivity." *Philosophical Studies* 172: 1243–63.

Costa, S., H. Lehmann, K. Dautenhahn, B. Robins, and F. Soares. 2015. "Using a Humanoid Robot to Elicit Body Awareness and Appropriate Physical Interaction in Children with Autism." *International Journal of Social Robotics* 7: 265–78.

Cramer, M. and C. Hauser. 2021. "Digidog, a Robotic Dog Used by the Police, Stirs Privacy Concerns." *New York Times*, February 27, 2021. https://www.nytimes.com/2021/02/27/nyregion/nypd-robot-dog.html.

Damiano, L. and P. Dumouchel. 2018. "Anthropomorphism in Human-Robot Co-Evolution." *Frontiers in Psychology* 9: 468.

Duffy, B. 2004. "Robots Social Embodiment in Autonomous Mobile Robotics." *International Journal of Advanced Robotic Systems* 1 (3).

Duffy, B. R., C. Rooney, G. M. P. O'Hare, and R. O'Donoghue. 1999. "What is a Social Robot?" Talk presented at the *10th Irish Conference on Artificial Intelligence & Cognitive Science, University College Cork, Ireland, 1–3 September, 1999*. Text accessed May, 2, 2023. http://hdl.handle.net/10197/4412.

Ekman, P. 1992. "An Argument for Basic Emotions". *Cognition and Emotion* 6: 169–200.

Ekman, P. and R. J. Davidson, eds. 1994. *The Nature of Emotion: Fundamental Questions*. Oxford University Press.

Epley N, A. Waytz, and J. T. Cacioppo. 2007. "On Seeing Human: A Three-Factor Theory of Anthropomorphism." *Psychological Review* 114 (4): 864–86.

Fong, T., I. Nourbakhsh, and K. Dautenhahn. 2003. "A Survey of Socially Interactive Robots: Concepts, Design, and Applications." *Robotics and Autonomous Systems* 42: 143–66.

Fussi, A. 2022. "Anger, Hatred, Prejudice. An Aristotelian Perspective." In *Empathy, Intersubjectivity, and the Social World: The Continued Relevance of Phenomenology. Essays in Honour of Dermot Moran*, edited by A. Bortolan and E. Magri, 389–412. De Gruyter.

Griffiths, P. E. and A. Scarantino. 2009. "Emotions in the wild. The situated perspective on emotion." In *The Cambridge Handbook of Situated Cognition*, edited by P. Robbins and A. Murat, 437–453. Cambridge University Press.

Happé, F. and U. Frith. 2006. "The Weak Coherence Account: Detail-Focused Cognitive Style in Autism Spectrum Disorders." *Journal of Autism and Developmental Disorders* 36: 5–25.

Hegel, F., C. Muhl, B. Wrede, M. Hielscher-Fastabend, and G. Sagerer. 2009. "Understanding Social Robots." In *2009 Second International Conferences on Advances in Computer-Human Interactions*, edited by S. Dascalu and I. Pupyrev, 169–74.

Henschel, A., G. Laban, and E. S. Cross. 2021. "What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You." *Current Robotics Report* 2: 9–19.

Hess, U. and S. Hareli. 2016. "The Impact of Context on the Perception of Emotions." In *The Expression of Emotion: Philosophical, Psychological and Legal Perspectives, edited by C. Abell and J. Smith*, 199–218. Cambridge University Press.

Hibberd, J. 2017. "Black Mirror Creator Explains that 'Metalhead' Robot Nightmare." *Entertainment Weekly*, December 29, 2017. https://ew.com/tv/2017/12/29/black-mirror-metalhead-interview/.

Huijnen, C. A. G. J., M. A. S. Lexis, and L. P. de Witte. 2016. "Matching Robot KASPAR to Autism Spectrum Disorder (ASD) Therapy and Educational Goals." *International Journal of Social Robotics* 8: 445–55.

ITV News. 2017. "Zeno: The Robot Teaching Autistic Children how to Communicate." 16 February, 2017. News report, 2:23. https://www.youtube.com/watch?v=8-q9tsZEoJs.

Kaliouby, R., R. Picard, and S. Baron-Cohen. 2006. "Affective Computing and Autism." *Annals of the New York Academy of Sciences* 1093: 228-48.

Krause-Jensen, K. and R. Rodogno. 2023. "Shame on Wrong Planet: A Study of Shame among People on the Autism Spectrum." In *The Moral Psychology of Shame*, edited by A. Fussi and R. Rodogno, 221–242. Rowman & Littlefield.

Krueger, J. 2014. "Affordances and the Musically Extended Mind." *Frontiers in Theoretical and Philosophical Psychology*, 4 (1003): 1–13.

Massantini, L. 2020. "Affective Scaffolds of Nostalgia." *Philosophical Inquiries* 8 (2): 47–70.

Meirav, A. 2009. "The Nature of Hope." *Ratio* 22 (2), 216–33.

Miceli, M. and C. Castelfranchi. 2010. "Hope: The Power of Wish and Possibility." *Theory and Psychology* 20 (2): 251–76.

Mori, M. 1970. "The Uncanny Valley." *Energy* 7: 33–5.

Naneva, S., M. Sarda Gou, and T. L. Webb. 2020. "A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust towards Social Robots." *International Journal of Social Robotics* 12: 1179–201.

Ocasio-Cortez, A. (@AOC). 2021. "Please ask yourself: when was the last time you saw next-generation, world class technology for education, healthcare, housing, etc consistently prioritized for underserved communities like this?" Twitter, February 25, 2021. https://twitter.com/AOC/status/1365023067769098245.

Russell, J. A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology* 39 (6): 1161–78.

Russell, J. A. and L. F. Barrett. 1999. "Core Affect, Prototypical Emotional Episodes and Other Things Called Emotions: Dissecting the Elephant." *Journal of Personality and Social Psychology* 76 (5): 805–19.

Šabanović, S. 2010. "Robots in Society, Society in Robots." *International Journal of Social Robotics* 2: 439–50.

Sartorato, F., L. Przybylowski, and D. K. Sarko. 2017. "Improving Therapeutic Outcomes in Autism Spectrum Disorders: Enhancing Social Communication and Sensory Processing Through the Use of Interactive Robots." *Journal of Psychiatric Research* 90: 1–11.

Schadenberg, B. R., D. K. J. Heylen, and V. Evers. 2018. "Affect Bursts to Constrain the Meaning of the Facial Expressions of the Humanoid Robot Zeno." In *Joint Proceedings of the Workshop on Social Interaction and Multimodal Expression for Socially Intelligent Robots and the Workshop on the Barriers of Social Robotics take-up by Society, edited by* C. Tsiourti, S. Hanke, and L. Santos, 30–9. CEUR.

Slaby, J. 2016. "Mind Invasion: Situated Affectivity and the Corporate Life Hack." *Frontiers in Psychology* 7.

Stephan, A., S. Walter, and W. Wilutzky. 2014. "Emotions beyond Brain and Body." *Philosophical Psychology* 27 (1): 65–81.

University of Herfordshire. n.d. "A Kaspar in every School." University of Hertfordshire. Accessed April 10, 2022. https://www.herts.ac.uk/kaspar/a-kaspar-in-every-school.

Urquiza-Hass, E. G. and K. Kortschal. 2015. "The Mind behind Anthropomorphic Thinking." *Animal Behaviour* 109: 167–76.

Walker, M. 2006. *Moral Repair. Reconstructing Moral Relations after Wrongdoing*. Cambridge University Press.

Winnicott, D. W. 1971. *Playing and Reality*. Routledge.

Zaveri, M. 2021. "N.Y.P.D.'s Robot Dog Returns to Work, Touching Off a Backlash." *New York Times*, April 14, 2021. https://www.nytimes.com/2021/04/14/nyregion/robot-dog-nypd.html.