PASSION

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

Editors in chief

Alfred Archer

Heidi Maibom

02

Table of content

Editor's Introduction: How Should We Think about Co-Occurring Emotions?

Heidi L. Maibom - University of the Basque Country/University of Cincinnati

Emotion theory tends to be concerned with emotions as if we ever only experienced one emotion at a time. It also tends to ignore, or at least downplay, emotions' temporal extension, on the assumption that our descriptions of them fit equally at all points of their unfolding. In other words, we are being told a story in which an emotion is a discrete and isolated event nicely delimited by the situations that give rise to it, on one side, and the behaviour or set of behaviours it tends to provoke, on the other. Of course, nobody denies that an emotion may take time to unfold, but the assumption seems to be that it stays largely the same throughout this period of time. This is an unjustified assumption, as the now extensive literature on emotion regulation shows (e.g. Gross 2014a). For we often modulate our emotions through various means. Most famously, we down-regulate emotions that feel bad, such as sadness, anger, or pain. Less researched, but no doubt equally important, is the *up-regulation* of positive emotions, such as joy or pride. What happens in these instances is that the emotion is modified by some, usually effortful, intervention on the subject's side. This might consist in re-evaluating the significance of its cause (cognitive re-evaluation), suppressing its expression, changing one's attentional focus, or modifying the eliciting situation (Gross 2014b). Most of these methods are taken right out of the Stoic playbook (see, e.g., "On Anger" in Seneca ca. 40/2020). Eastern religions, such as Tibetan Buddhism, have other suggestions for a therapy of emotions, such as making oneself experience "an opposite" emotion (McRay 2015).

Now, it might be thought that we modify our emotions under more-or-less extraordinary circumstances, during which we exert real effort, so that, most of the time, they run their usual course in orderly fashion. This is belied by the following facts. People working on emotion regulation find it useful to stress that regulation can be either conscious or unconscious (e.g., Eisenberg, Hofer, Sulik, and Spinrad 2014), suggesting that it is much more common than suggested above. Psychiatrists insist that emotion regulation is central to mental health (e.g. Berking and Wupperman 2012; Menefee 2022), and some empathy researchers have argued that good emotion regulation is needed for empathy to lead to prosocial action (Spinrad and Eisenberg 2014). Moreover, Jamil Zaki (2014) has collected an impressive amount of data indicating that empathy is motivated, and that people use a variety of means to *avoid* or *reduce* the experience of empathy for those who suffer. Additionally, there is much evidence that doctors and nurses downregulate their empathic response to pain more or less automatically (Agledahl et al. 2011; Gleichgerrcht and Decety 2014; Hunt, Denieffe, and Gooney 2017; Neumann et al. 2011).

But what this issue focuses on is not the temporal dimension of emotions as such, but on how isolated they actually are. This question turns out not to be altogether separate from that concerning emotions' temporal dimension. Why? Because an overall emotional experience appears to be a composite result of what other emotions are experienced at the same time, or during intervals in the unfolding of an emotion. For instance, empathic distress and sympathy are almost invariably experienced at the same time when someone is presented with a person in need (Batson, Early, and Salvarini 1997; Carrera et al. 2013; Grynberg and López-Pérez 2018). Nonetheless, they are characterised as quite different emotions, not just in their affective qualities, but also with respect to their motivational effects. What happens at the level of psychological theorising with respect to empathic distress and sympathy is quite similar to the treatment of shame and guilt in the psychological literature. Essentially, one emotion is charged with all, or as much as possible of, the negatively valenced, morally and behaviourally problematic aspects, and the other is charged with the positive. Thus, empathic distress is typically described as *personal* and as leading to egoistic motivation, even though the evidence we do have suggests that people who experience distress at another's distress experience an equal amount of distress for the subject and for themselves (Batson, Early, and Salvarini 1997). Sympathy is the warm-hearted and concerned counterpoint to the worried and stressful affect just described, and can be morally and behaviourally beneficial; it leads to helping the person in distress, which is often the morally right thing to do.

When it comes to guilt and shame, the situation is much the same. These are emotions that are very often experienced at the same time, and which are difficult to pry apart (see, e.g., Tangney et al. 1996). Hence the many papers offering a distinction, particularly among philosophers. Psychologists, on the other hand, have settled, more or less uncritically, on the following. Shame is a problematic emotion affecting a person's sense of their own value, and guilt is a prosocial emotion concerning an agent's *actions.* Consequently, shame is operationalised as an almost exclusively *negative* emotion in the most commonly used measures of guilt and shame, TOSCA (Tangney, Wagner, Dearing, and Gramzow 2000; for a critique see, e.g., Maibom 2019). Guilt, by contrast, contains all of the positive aspects of the emotional episode, namely in motivating the person to repair and in being less aversive than shame. We therefore have lots of papers showing that shame is highly problematic and guilt very good. But this could not be otherwise, given the measure used. I don't need to tell you why that is problematic. When shame is more positively construed —and not as an extremely strong aversive emotion, which leads to retreat or reactive aggression—it turns out that the psychological data suggests that it leads to some of the more positive behaviours that studies using TOSCA would only ascribe to 'guilt' (Berndsen and McGarty 2012; Berndsen and Gausel 2015; Fessler 2004; De Hooge, Breugelmans, and Zeelenberg 2008; De Hooge, Zeelenberg, and Breugelmans 2010).

All of this strongly suggests that either the boundary between shame and guilt, on the one hand, and empathic distress and sympathy, on the other, is much more fluid than current characterisations lead us to believe (*and* that we need to be very careful in philosophising about such emotions not to rely uncritically on the psychological research). Recognising this interesting fact encourages us to ask new questions, such as "What effect does the fact that I experience *both* guilt and shame have for each of these emotions?" "Does their co-occurrence *alter* how each are experienced, how the subject comes to think about the relevant action or herself, and what effect, if any, does it have for any subsequent motivation?" Or "Is guilt felt with shame just the same as guilt felt without shame?"

One reason to think that co-experienced emotions might have unique effects of their own is that it seems that if people experience just sympathy for someone in need, but little or no distress at their distressing situation or their distressed feelings, they are less likely to help that person (Lockwood, Seara-Cardoso, and Viding 2014;

Cameron and Payne 2011). Conversely, people who experience more distress alongside their sympathy are *more* likely to help people in need (Cameron et al. 2019; Barraza and Zak 2009). When that distress reaches a certain level, however, helping behaviour starts falling off and more selfish action increases if it is easier than helping (Batson 2011). It is, therefore, not at all unlikely that the prosocial effects that we are trying to ascribe to *one* emotion is, in fact, the result of the experience of *two* emotions. Moreover, some evidence suggests that the way the two interact is also rather consequential for subsequent motivations. Distress can be experienced quite strongly in the beginning or the middle of an empathic episode, but if it is the predominant emotion towards the end, the chances of the agent helping the person in need plummets. If, on the other hand, sympathy has the upper hand towards the end, then there are very good chances that the person will help (Carrera et al. 2013).

This second issue of *Passion* addresses the issue of how to think of emotions in light of their co-occurrence. The issue is (mostly) the final product of talks given at an EPSSE pre-conference workshop in Graz, which I organised in June 2022. We start with an article by Christiana Werner, who argues that the existence of concurrent emotions—emotions experienced simultaneously—complicates the task for empathisers. Typically, people tend to just empathise with one emotion. But people often experience different—sometimes apparently opposing—emotions. She gives the example of Desdemona from Jeffrey Eugenides's *Middlesex*, who, upon finding her husband dead on the kitchen floor, experiences both immeasurable grief *and* relief. This challenges the idea that we can generally reach phenomenal knowledge of others through empathy, since the reproduction of such emotions would be a very difficult task indeed. Instead, Werner argues, it is really phenomenal *understanding* we aim for in empathy, i.e., an understanding that is *close enough*.

Werner's paper is followed by one by Edward Harcourt, who wrestles with the distinction between shame and guilt and their respective connections with "the morality system" (following Bernard Williams 1985). As he notes, the difference between the two is much exaggerated. Nonetheless, he believes that, whereas guilt has a connection with the morality system, shame, when properly analysed, does not.

Laura Candiotto engages deeply with the diaries and letters of Etty Hillesum, a Dutch-Jewish woman who eventually perished in the camps. What is special about this woman is her ability to feel contentment at her suffering with other victims of the Nazi terror. We can see the willingness to feel for others in awful situations as being ennobling, Candiotto insists. Her claim is well supported by the empathy literature, which suggests that people spend a good amount of effort trying to *avoid* experiencing empathy for those who suffer (e.g. Zaki 2014), and when they do manage to do so, they are less willing to help those in need (Cameron, Harris, and Payne 2016; Cameron and Payne 2011). It is therefore plausible that *choosing* to experience the suffering of others *is* moral progress of sorts, assuming that it leads to something other than pure suffering. And here the combination of contentment and empathic distress is very promising. For one of the standard arguments one sees *against* empathising with people who suffer is that it just creates more suffering and might lead to bad psychological outcomes (see, e.g., Bloom 2016). However, an ability to be content in one's empathic or compassionate relating to others is yet another example of how taking concurrent emotions into account can make a real difference to theorising about emotions.

Aleksandra Hernandez uses reactions to contrasting or conflicting emotions as a mirror to what she calls "the epistemic landscape." Those whom the landscape favours, she argues, quickly resolve whatever emotional conflicts arise and are in any case prone to experiencing emotions that prevent them from re-examining the presuppositions of this landscape, which is in their favour. Oppressed groups, such as women and other factual minorities, on the other hand are more likely to experience epistemic anxiety as a result of these contrasting

emotions, and that, in turn, makes them better knowers. The way this unlevel epistemic playing field works in practice is illustrated by an analysis of Joy Williams' short story *Shepherd*.

Katharina Anna Sodoma examines co-occurring emotions through the lens of what she calls meta-emotions. These are emotions directed at other emotions. For instance, we might be embarrassed about loving a schlocky movie, angry that we are ashamed about being different, or happy that we are finally able to feel the sadness that we have bottled up for so long. Experiences such as these are very common, Sodoma insists, but have not received a lot of attention by emotion researchers so far. Her focus is specifically on affective empathy. Here, these emotions create challenges for the empathiser when it is the target who experiences them. More interesting, perhaps, are the ways in which the empathiser's emotional reactions to her empathic affect opens up for rather profound self-reflection and other-directed assessment.

Finally, Ditte Marie Munch-Jurisic argues against the idea that we are typically aware of what we feel and why we feel it. Instead, she presents a persuasive case for emotional "disorientation" being the norms. This is exemplified by a common observation that we are often able to explain our emotions in quite disparate ways and that those ways often change with time as we find new interpretations better suited for our current projects and ideas. Like many others in this issue, Munch-Jurisic uses empathy as an example. She critiques Dan Batson's framework by showing how it fails to account for a soldier who's suffered moral injury.

I hope you find this issue intriguing and that it will provoke you to think more about how co-occurrent emotions should change the way we think about emotions.

References

Agledahl, K. M., P. Guldbrandsen, R. Førde, and Å. Wifstad. 2011. "Courteous but not Curious: How Doctors' Politeness Masks their Existential Neglect. A Qualitative Study of Video-Recorded Patient Consultations." *Journal of Medical Ethics* 37: 650–4.

Barraza, J. A. and P. J. Zak. 2009. "Empathy towards Strangers Triggers Oxytocin Release and Subsequent Generosity." *Annals of the New York Academy of Sciences* 1167: 182–9.

Batson, C. D. 2011. *Altruism in Humans.* Oxford University Press.

Batson, D. C., S. Early, and G. Salvarini. 1997. "Perspective Taking: Imagining How Another Feels versus Imagining How You Would Feel." *Personality and Social Psychology* 23: 751–58.

Berking, M. and P. Wupperman. 2012. "Emotion Regulation and Mental Health: Recent Findings, Current Challenges, and Future Directions." *Current Opinions in Psychiatry* 25: 128–34.

Berndsen, M. and C. McGarty. 2012. "Perspective Taking and Opinions about Forms of Reparation for Victims of Historical Harm." *Personality and Social Psychology Bulletin* 38: 1316–28.

Berndsen, M. and N. Gausel. 2015. "When Majority Members Exclude Ethnic Minorities: The Impact of Shame on the Desire to Object to Immoral Acts." *European Journal of Psychology* 45: 728–41.

Bloom, P. 2016. *Against Empathy.* Ecco.

Carrera, P., L. Oceja, A. Caballero, D. Muñoz, B. López-Pérez, and T. Ambrona. 2013. "I Feel So Sorry! Tapping the Joint Influence of Empathy and Personal Distress on Helping Behavior." *Motivation and Emotion* 37: 335–45.

Cameron, C. D., L. T. Harris, and B. K. Payne. 2016. "The Emotional Cost of Humanity: Anticipated Exhaustion Motivates Dehumanizing of Stigmatized Targets." *Social Psychological and Personality Science* 7: 105–12.

Cameron, C. D. and B. K. Payne. 2011. "Escaping Affect: How Motivated Emotion Regulation Creates Insensitivity to Mass Suffering." *Journal of Personality and Social Psychology* 100: 1–15.

De Hooge, I. E., S. M. Breugelmans, and M. Zeelenberg. 2008. "Not So Ugly after All: When Shame Acts as a Commitment Device." *Journal of Personality and Social Psychology* 95: 933–43.

De Hooge, I. E., M. Zeelenberg, and S. M. Breugelmans. 2010. "Restore and Protect Motivations following Shame." *Cognition and Emotion* 24: 111–27.

Eisenberg, N., C. Hofer, M. J. Sulik, and T. L. Spinrad. 2014. "Self-Regulation, Effortful Control, and their Socioemotional Correlates." In *Handbook of Emotion Regulation*, 2nd ed., edited by J. J. Gross, 156–71. The Guilford Press,.

Fessler, D. M. 2004. "Shame in Two Cultures: Implications for Evolutionary Approaches." *Journal of Cognition and Culture* 4: 207–62.

Gleichgerrcht, E. and J. Decety. 2014. "The Relationship between Different Facts of Empathy, Pain Perception and Compassion Fatigue among Physicians." *Fronters in Behavioral Neuroscience* 8: 243.

Gross, J. J., ed. 2014a. *Handbook of Emotion Regulation*, 2nd ed. The Guilford Press.

Gross, J. J. 2014b. "Emotion Regulation: Conceptual and Empirical Foundations." In *Handbook of Emotion Regulation*, 2nd ed., edited by J. J. Gross, 3–20. The Guilford Press.

Grynberg, D. and B. López-Pérez. 2018. "Facing Others' Misfortune: Personal Distress Mediates the Association between Maladaptive Emotion Regulation and Social Avoidance." *PLOS One.*

Hunt, P. A., S. Denieffe, and M. Gooney. 2017. "Burnout and its Relationship to Empathy in Nursing: A Review of the Literature." *Journal of Research in Nursing* 22: 7–22.

Maibom, H. L. 2019. "Empathy and Emotion Regulation." *Philosophical Topics* 47: 149–63.

McRae, E. 2015. "Buddhist Therapies of the Emotions and the Psychology of Moral Improvement." *History of Philosophy Quarterly* 32 (2): 101–22.

Menefee, D. S. 2022. "The Importance of Emotional Regulation in Mental Health." *American Journal of Lifestyle Medicine* 16: 28–31.

Neumann, M., F. Edelhaeuser, D. Tauschel, M. R. Fischer, M. Wirtz, C. Woopen, A. Haramati, and C. Scheffer. 2011. "Empathy Decline and its Reasons: A Systematic Review of Studies with Medical Students and Residents." *Academic Medicine* 86: 996–1009.

Seneca, L.A. ca. 40/2020. *On Anger.* In: Lucius Annaeus Seneca: *Anger, Mercy, Revenge.* R.A. Kaster & M. Nussbaum (Transl.). Chicago: University of Chicago Press.

Spinrad, T. and N. Eisenberg. 2014. "Empathy and Morality: A Developmental Psychological Perspective." In *Empathy and Morality*, edited by H. L. Maibom, 59–70. Oxford University Press,.

Tangney, J. P., P. E. Wagner, R. L. Dearing, and R. Gramzow. 2000. *The Test of Self-Conscious Affect-3 (TOSCA-3).* APA PsycTests.

Williams, B. 1985. *Ethics and the Limits of Philosophy.* Fontana.

Zaki, J. 2014. "Motivated Empathy." *Psychological Bulletin* 140: 1608–47.

Concurrent Emotions, Affective Empathy, and Phenomenal Understanding

Christiana Werner - Universität Duisburg-Essen

Abstract

According to an optimistic view, affective empathy is a route to knowledge of what it is like to be in a target person's state ("phenomenal knowledge"). Roughly, the idea is that the empathiser gains this knowledge by means of empathically experiencing the target's emotional state. The literature on affective empathy, however, often draws a simplified picture according to which the target feels only a single emotion at a time. Co-occurring emotions ("concurrent emotions") are rarely considered. This is problematic, because concurrent emotions seem to support a sceptical view according to which we cannot gain phenomenal knowledge of the target person's state by means of affective empathy. The sceptic concludes that attaining the epistemic goal of affective empathy is difficult, in practice often impossible. I accept the sceptic's premises, but reject the conclusion, because of the argument's unjustified, hidden premise: that the epistemic goal of affective empathy is phenomenal knowledge. I argue that the epistemic goal of affective empathy is phenomenal understanding, not knowledge. Attention to the under-explored phenomenon of concurrent emotions clarifies why this is important. I argue that this is the decisive epistemic progress in everyday cases of phenomenal understanding of another person.

Keywords: Emotions, Empathy, Affective Empathy, Interpersonal Understanding, Phenomenal Concepts

1. Introduction

According to an optimistic view, empathy is a route to knowledge of other people's mental states. This could be propositional knowledge. Many philosophers also claim that empathy can be a route to knowledge of what it is like to be in a target person's state. This article focusses on the question of whether empathy can be a source of the latter type of knowledge. Roughly, the idea is that the empathiser gains this knowledge by means of empathically co-experiencing the target person's emotional state.

In a first step, I identify affective empathy as an obvious candidate of a mental process that provides knowledge of what it is like to be in the target's situation. I understand knowledge of what it is like as knowledge of the phenomenal character of conscious states and use the term "phenomenal knowledge."

In a second step, I show that the literature on affective empathy, however, often draws a simplified picture according to which the target feels only a single emotion at a time. Co-occurring emotions (labelled here as "concurrent emotions" in order to avoid confusion between the emotions a subject feels simultaneously, on the one hand, and emotions which are empathically co-experienced, on the other hand) are rarely considered. This is problematic, because concurrent emotions seem to support a sceptical view according to which we cannot gain phenomenal knowledge of the target person's state by means of affective empathy.

The sceptic could argue as follows: first, the phenomenal character of a single mental state, like that of fear, is different from the phenomenal character of an overall mental state comprised of more than one emotion. Plausibly, the latter is not merely the additive result of the respective concurrent emotions, but involves their reciprocal modification. I will introduce the notion of an "atomic conscious state" for a single mental state, and a "complex conscious state" for a person's overall mental state, comprised of a set of simultaneous states, some of which are emotional states. Second, if we want to know what it is like for the target to be in her situation, it is phenomenal knowledge of her complex conscious state that we are after. Third, although not metaphysically impossible, it is psychologically far more difficult to empathically experience the target's complex conscious state than just a single emotion. The sceptic concludes that attaining the epistemic goal of empathy is difficult, in practice often impossible.

Ultimately, I accept the sceptic's premises, but reject the conclusion, because of the argument's unjustified, hidden premise: that the epistemic goal of affective empathy is phenomenal knowledge. I argue that the epistemic goal of affective empathy is phenomenal *understanding*, not knowledge. Attention to the under-explored phenomenon of concurrent emotions clarifies why this is important. The phenomenal complexity of the target's complex conscious state is one important reason why affective empathy comes in degrees: the empathiser's empathic experience can match the target's complex conscious state more or less precisely. Even if the empathizer cannot recognise and empathically experience the full range of the target's concurrent emotions, the empathiser can gain selective phenomenal knowledge or phenomenal beliefs that are close to the truth. Here phenomenal understanding comes into play, which, in contrast to phenomenal knowledge, is not factive and, like affective empathy, comes in degrees. Hence, understanding the target on the basis of selective phenomenal knowledge, and even on the basis of false phenomenal beliefs, is possible, if they are close enough to the truth. This is epistemic progress, because the empathiser is in an epistemically better position than a person who completely fails to empathise with the target, or makes no attempt to empathise. I argue that this is the decisive epistemic progress in everyday cases of phenomenal understanding of another person.

2. Empathy

Empathy is often spelled out in terms of simulation. The idea is, very roughly, that the empathiser engages in a process of simulating the target person's mental state. It is very common to distinguish between so-called cognitive and affective empathy (see e.g. Batson 2009, Hoffmann 2011, Kaupinnen 2017, Maibom 2017, Spaulding 2017). These terms can be misleading in at least two ways. First, they are misleading because both the target and the empathiser herself can play host to both cognitive and affective states. I will understand the

latter term as picking out qualifications of the empathic process. Second, the terms can be misleading because it is not the case that affective empathy is a process without any cognitive elements. Thus, I understand cognitive empathy as a process of non-affectively, merely propositionally imagining or simulating the target's state. Affective empathy, in contrast, is a process during which the empathiser simulates the target's states in such a way that she feels or co-experiences the target's affective states. This process can of course also involve cognitive elements.

Empathy, understood as a process of mental state recognition, has an epistemic value. One idea is that by means of empathy an empathiser learns the target person's mental state. This knowledge can be spelled out in terms of propositional knowledge. There is, however, also the idea that empathy is a way of gaining knowledge of what it is like ("phenomenal knowledge") for the target to be in her state. According to Steinberg, for example, empathy is "the affective apprehension of the mental state of another, a way of gaining insight into what it is like to be another person." Some philosophers claim that knowledge of what it feels like, or how one feels, is even the primary epistemic function of empathy (see Smith 2017 and for discussion Stueber forthcoming). We find the idea that empathy is a way of gaining knowledge of what it is like for the target of empathy to be in her specific state in various places in the literature (for example Coplan 2011, Boisserie-Lacroix and Inchingolo 2021, Wiltsher 2021).

It seems that affective empathy is the obvious candidate we should start with in trying to find out whether empathy can be a source or way of gaining knowledge of what it is like for somebody else to be in a specific mental state. Because emotions play an important role in a person's psychology, philosophers are mainly interested in empathy as a source of phenomenal knowledge of another person's emotional or affective states. The question is then whether affective empathy can be a source of phenomenal knowledge of the target person's emotional states.

It seems that affective empathy could be a source of phenomenal knowledge, if the empathiser simulates the target's emotional state so that the empathiser herself co-experiences this state and thereby learns what it is like to be in that state.

In the following I provide several examples which suggest that the target's emotional state is understood as a single emotion, rather than concurrent emotions. Heidi Maibom gives this definition:[1]

> Person S empathizes with person O's experience of emotion E in situation C if S feels E for O as a result of believing or perceiving that O feels E, or imagining being in C. (Maibom 2017, 22)

Whereas this definition leaves open whether the underlying process or state is imaginative, perceptive, or a (set of) beliefs, the word "feel" indicates that the empathiser herself is in an affective state. Thus, this empathic process is not a mere a state of "cold" recognition of the target's state. According to Maibom, the empathiser feels the same emotion "E" as the target.

Amy Coplan names an affective matching whose formulation indicates she also thinks of a single emotion as the target's state: "The empathizer must therefore experience the same type of emotion (or affect) as the target." (Coplan 2011, 6).

---

1   Later in this paper Maibom (23–24) specifies that the empathiser's experience can be an emotion similar to the one the target's experiences, and need not be exactly the same.

Similarly, Boisserie-Lacroix and Inchingolo argue for the epistemic significance of knowledge of what it is like to be in the target's state (they use the term "phenomenal insight") in empathically understanding other people. They accordingly hold an optimistic view on empathy as a route to this type of knowledge:

> It is because we are aware of what it feels like to be in the grip of an emotion that our empathizing will be successful. . . . It is because the interpreter can imagine what it feels like to be in a state of anger that she is able to understand why Jane tore Joan's photo. Empathizing with Jane provides her with a phenomenal insight into the way anger can motivate one to act arationally. (Boisserie-Lacroix and Inchingolo 2021, 7107)

And Olivia Bailey writes:

> In certain critical respects, the emotional experience of the one who empathizes closely resembles the emotional experience of the target of empathy. . . . The widower apprehends his loss through the lens of grief. We as the widower's empathizers also allow our thoughts to be directed in the ways characteristic of grief . . . the isomorphism between this empathetic experience and the original grief of the widower strongly recommends the conclusion that when we empathize, we do not merely imagine that we are feeling some emotion. Rather, we actually experience an emotion. (Bailey 2022, 4)

All these quotes support in slightly different ways the idea that there is a process of co-experiencing the target's state. For this reason, such a process is a good candidate for a source of phenomenal knowledge of another person's affective state. However, these quotes also support the supposition that many philosophers in the debate about empathy consider only a single emotional state of the target person.

Before we can proceed to the problem of concurrent emotions and investigate whether empathy can be a source of phenomenal knowledge of others' affective states, we need to specify this type of knowledge.

3. Phenomenal Knowledge and Phenomenal Concepts

There are different physicalist and dualist approaches to explaining the peculiar sort of knowledge we gain exclusively from experience. For our purposes here, we don't need to assume that it is for metaphysical reasons that it is not possible to have phenomenal knowledge of a conscious state we have not yet experienced. Instead, a much weaker claim is sufficient: although it might be metaphysically possible to gain phenomenal knowledge in other ways than by means of experience, in practice, people have to experience a conscious state in order to gain phenomenal knowledge of this state.

According to a widely used strategy, the epistemic progress provided by experience is best explained in terms of phenomenal concepts. Experiencing a certain state with phenomenal qualities for the first time generally enables the acquisition of the phenomenal concept of the state in question. At least in practice, we only gain a phenomenal concept of a certain state by experiencing that state. Hence, we have no phenomenal concept of a state we have so far not experienced. This will be true even when the differences between the phenomenal character of some state we have already experienced and the unexperienced state are fairly minimal.

The deployment of phenomenal concepts allows us to explain the difference between phenomenal and non-phenomenal beliefs. We find the idea of this distinction expressed by Peter Goldie (2002), who describes the thought experiment of Irene, the icy-cool ice-scientist, a thought experiment that concerns the phenomenal qualities not of perception, but of the emotional state of fear. Irene experiences fear for the first time when she slips and falls on the ice. Goldie argues:

> Before, when she thought of someone being afraid, her thought was restricted, roughly, to the causal role that that person's experience played—its typical causes and effect; and this way of thinking is one that might be available to, say, a Martian who was incapable of feelings of fear. Now, when she judges that someone else is afraid, she can deploy in the thought her newly gained phenomenal concept of fear. (Goldie 2002, 246)

According to many philosophers of mind (for example Papineau 2002), we need to distinguish psychological concepts from phenomenal concepts. Both are expressed through everyday terms for psychological states such as pain. Phenomenal concepts can be deployed in different ways, such as in a propositional structured belief.

Following Papineau, we can say that, prior to her first fear experience, Goldie's Irene had a psychological concept of fear, but no phenomenal one. Assuming that Irene knew the facts about fear that most of us know, such as its typical facial expressions, behaviours, and fear-inducing situations, we can also assume that she was able to recognise and believe that another person was afraid. This means that Irene had a concept of fear, but it was a mere psychological concept. Before her own experience of fear, she had no phenomenal concept of fear to deploy in, for example, believing that someone was afraid. Once she has gained the phenomenal concept and recognised that her experience was an experience of fear, she is able to deploy the phenomenal concept so that she can phenomenally believe that someone is in a mental state such as being afraid (see Nida-Rümelin 1998).

We are now able to further specify why having a phenomenal belief about another person's emotional state is epistemically better than just non-phenomenally knowing that the person is in that state: having a phenomenal concept is an epistemic benefit, because it is only then that we are able to recognize that, e.g., a mental state falls under both a physical and a psychological concept and a specific phenomenal concept, or compare a phenomenal concept with another one.[2]

4. Atomic and Complex Conscious States

In this section I will introduce a distinction between atomic conscious states and complex conscious states. This distinction will help to clarify what an empathiser can simulate within the process of affective empathy. In order to do so, we need to draw attention to details of Frank Jackson's seminal thought experiment in

---

2 A proper account of phenomenal belief and phenomenal concepts would have to say more about how phenomenal concepts are integrated into propositions. This task in turn requires further specification of phenomenal concepts' nature. There are a number of different accounts of phenomenal concepts (for overviews see Balog 2009, Sundström 2011), which have different implications for how to conceive phenomenal beliefs. Such accounts have to answer various questions: what, for example, is the precise relation of the phenomenal concept to its referent? And does a thought with a content which includes a phenomenal concept necessarily involve the relevant experience, as is suggested by quotational accounts of phenomenal concepts (Papineau 1993, chap. 4; 2002, chap. 4; 2007; Balog 1999, Melnyk 2002), or is the connection to the experience better conceived in dispositional terms? In this paper, I have to leave these and other related question unanswered.

formulating the so-called knowledge argument. In the thought experiment, the scientist Mary knows all physical facts about colours and colour perception. But she has spent her whole life trapped in a black-and-white room and thus has never seen any colours (Jackson 1982). Jackson claims that Mary learns something when she leaves her room and perceives red for the first time, namely what it is like to see red. Since she knows all physical facts, the argument goes, Mary learns a non-physical fact about colours or colour perception. Thus, there are non-physical facts.

As mentioned in previous section, the metaphysical dimensions of the thought experiment are not central for the aim of this paper. It is, however, important to draw attention to one feature of the thought experiment and its discussion: Jackson's example of a conscious state, and many other examples we find in the literature on the knowledge argument, are of colour perceptions, as well as some other sensual experiences or states of pain.

These examples are of conscious states which can be distinguished by their phenomenal character. For the purposes of this paper, I assume that there are psychological states with a unique phenomenal character, such as colour perceptions or other perceptual states. Their character has no further phenomenal components or they are "undifferentiated wholes" (Kind 2020, 144). For this reason, I call these states "atomic conscious states." I assume that not only colour perceptions, but also many affective states, are atomic states in this sense. In Jackson's thought experiment, the scientist Mary gains knowledge of such an atomic state. In terms of phenomenal concepts, Mary gains – in addition to the physical concept "red," which she already possessed – the phenomenal concept "red."

Most of the time, we are not only in one single atomic conscious state. For example, we do not only perceive the colour red, but perceive several objects with shapes and different colours. Beside this visual experience, one could have other sensual experiences, such as hearing a song from the radio and smelling the flowers in the vase on the desk. And besides these sensual experiences, a person can have occurrent beliefs and desires, all more or less at the same time. That means that people often—or even normally—are in a conscious state that is multi-layered.[3] In the following I will refer to this conscious total temporary state by means of the term "complex conscious state."

5. Concurrent Emotions

The aim of this section is to see how emotions fit into this picture drawn in terms of a distinction between atomic and conscious states.

Prima facie, it is an open question whether emotions are atomic states in the sense mentioned above at all. Different theories of the emotions answer the question about the ontological nature of the emotions differently. So-called belief-desire theories of the emotions (Green 1992) suggest that emotions are not mental states *sui generis*, but have to be analysed in terms of beliefs and desires (see also Marks 1982, Searle 1983). In contrast, more recent theories of the emotions claim that emotions are mental states *sui generis* (see for example Roberts 2003, Prinz 2004, Döring 2007, Tappolet 2016). If the assumption of these theories—that emotions are mental states sui generis—is true, emotions are atomic conscious states. For the sake of simplicity, I will work with this assumption in the following.

---

3   Grice uses "the same total temporary state" or "t.t.s." to refer to the simultaneous state of a person at a given time (Grice 1941, 88).

We usually experience an emotion whilst playing host to other mental states, for example beliefs and desires, which can intensify or attenuate the phenomenal character of that emotion. If I am afraid of a dog in front of me, the belief that it is leashed can attenuate the phenomenal quality of my fear, and the observation that it is suddenly free can have the opposite result. The phenomenal quality of the happiness a child feels on the last day before the summer holidays can be intensified because she hears that the weather will be sunny for the next couple of weeks, or attenuated if the opposite is the case.

It is important to acknowledge that emotions occur in conjunction with other emotions, too. We often find this described in literature. Here is an example taken from the novel *Middlesex* by Jeffrey Eugenides, where the protagonist's grandmother finds her husband (who is also her brother) dead on the kitchen floor:

> Desdemona had found Lefty on our kitchen floor, lying next to his overturned coffee cup. She knelt beside him and pressed an ear to his chest. When she heard no heartbeat, she cried out his name. Her wail echoed off the kitchen's hard surface: the toaster, the oven, the refrigerator. Finally she collapsed on his chest. In the silence that followed, however, Desdemona felt a strange emotion rising inside her. It spread in the space between her panic and grief. It was like a gas inflating her. Soon her eyes snapped open as she recognized the emotion: it was happiness. Tears were running down her face, she was already berating God for taking her husband from her, but on the other side of these proper emotions was an altogether improper relief. The worst had happened. This was it; the worst thing. For the first time in her life my grandmother had nothing to worry about. (Eugenides 2002, 244–45)

Eugenides describes the mental state of Desdemona in this terrible situation with great richness. In my terminology, Desdemona's complex state has several emotional states as components: panic, grief, happiness, and relief. The appearance of happiness is described in a remarkable way: Desdemona feels an additional emotion beside panic and grief, but it takes her a while to recognise which emotion this is. Although it is not mentioned in Eugenides' description, readers could assume that surprise could be another emotion Desdemona feels when she realises that there is a positive emotion among the negative ones.

Bayne and Chalmers claim that if a person is in more than one conscious state simultaneously, there is something it is like (Bayne and Chalmers 2003) for the person to be in these states simultaneously. The phenomenal quality of the experience of two or more concurrent conscious states is different from the phenomenal quality of each state individually. It seems hard to deny that the experience of being, for example, tired and hungry, is different from being tired and in pain. We can assume that we cannot only have phenomenal concepts of atomic conscious states, but also of complex conscious states.

The claim that there is something it is like to be in a complex conscious state leaves the question open of how the phenomenal character of this complex state is constituted. We could think of the phenomenal character of the components as simply added to each other in line with a building block model. According to this model, the phenomenal character of the complex state would be merely the sum of its components.

However, I think this view is oversimplified. Concurrent mental states can plausibly reciprocally influence each other's phenomenal character, for example in their intensity. Some conscious states may be foregrounded, others more in the background. This structural feature will have an influence on the overall experienced state. Therefore, the composition of the components is, or at least can be, more than the sum of its components.

Desdemona's complex state in this situation has, then, several emotional states as components. The phenomenal character of her complex state must be intense and characterised by the fact that her concurrent emotions mix opposed hedonic tone and valence: panic and grief are emotions with negative or unpleasant hedonic tone, in contrast to happiness and relief, which are both positive or pleasant.

We can think of other examples of concurrent emotions which pull in different directions and display mixed valences: we can pity someone who had a minor accident, whilst simultaneously feeling amused by the slapstick of the event; we can feel bored in a waiting room and at the same time angry because we have to wait; we can be afraid of a scary looking dog, but also feel pity for the animal because we can see that it has to live under terrible conditions[4].

Simultaneously experienced atomic states can mutually influence their phenomenal character. This can be observed, for example, in colour perception: one colour can take on different appearances when it is surrounded by different colours. A similar phenomenon occurs when we ordinarily perceive a square as white, but when it is surrounded by black squares it appears grey to us. Similar changes happen in experiences of other sensual states, such as taste and smell.

I assume that concurrent affective states can also influence the phenomenal character of an emotion. In the same way that a colour appears to have a different tone when perceived in differently coloured surroundings, the phenomenal character of an affective state can change. There can be an influence on the affective state's intensity, but also on its hedonic or otherwise felt tone. Anger, for example, can be experienced as less sharp when one is at the same time slightly amused[5].

There are phenomena which we experience as a whole, but which seem to be partly constituted by concurrent emotions, as in the case of nostalgia. Jesse J. Prinz (2004) explains nostalgia's bittersweet character through the simultaneous experience of happiness and sadness.

Moreover, there is most likely a difference in your overall complex experience if you are, in the one case, besides other non-affective states simply relieved about something or, in another, both relieved and worried about something else.

As we have seen that concurrent emotions can mutually influence and determine the phenomenal character of a complex conscious state of a person, we have arrived at a stage where we can discuss the influence of concurrent emotions on our ability to gain phenomenal knowledge of another person's state by means of affective empathy.

6. Affective Empathy and the Affective Matching Condition

So far we have taken note of the prominent idea that affective empathy is a source of phenomenal knowledge of a target person's mental, and in particular emotional, state. We have specified that a subject can acquire a phenomenal concept by means of experiencing a corresponding conscious mental state. In practice, people only gain phenomenal concepts of conscious states by means of experiencing them. Hence, we have only

---

4   For an overview in psychological literature on mixed emotions see Larson 2017.
5   For more examples and empirical data see, for example, Heavey et al. 2017, Roseman 2017.

phenomenal concepts of states we have experienced and no phenomenal concepts of states we have so far not experienced.

Since it is highly unlikely that any empathiser has been in the same complex conscious state as her target—including all its conscious components, such as concurrent emotions, desires, beliefs and sensual states—we can now see the supposed role of affective empathy more clearly: if an empathiser simulates the target's complex state correctly, she can then acquire the phenomenal concept of the complex state. In a second step, she could entertain a thought about the target's complex state by deploying the correct phenomenal concept. This shows why the empathiser needs to simulate the target's exact experience. It is only then that the empathiser can gain the correct phenomenal concept and, as a result, can form true phenomenal beliefs about the target's state.

Some authors argue that affective congruence between the target's emotion and the empathic emotion is sufficient for successful empathy. This means that there need only be a qualitative similarity between the target's affect state and the empathiser's state. If, for example, the target person is in an emotional state with a positive hedonic tone and valence, the empathic emotion needs to be of a positive hedonic tone and valence as well (Hoffmann 2000). Others argue for a stronger condition. According to the quotation from Maibom presented in the previous section, affective empathy is successful only if the empathiser feels the same emotion as the target person (Maibom 2017, 22). Amy Coplan argues similarly, but distinguishes between emotion type and the intensity of the experiential character of the emotion (Coplan 2011, 6).

It is important to note that such conditions concern the question of whether a mental process counts as (successful) empathy. The condition of affective congruence, for example, would not guarantee that empathiser and target feel the same. Hence, empathy with affective congruence would (typically) not be a source of phenomenal knowledge about the target person's state. For this reason, conditions for successful empathy have to be distinguished from a condition which guarantees that the mental mechanism named is a source of phenomenal knowledge of the target's state.[6]

In order to know how another person feels, we need to have phenomenal knowledge of the phenomenal character of her state. If successfully empathising is a way of experiencing the same emotional state as that of a target person, there has to be an identity of phenomenal character between the two. Type identity between the empathic emotional state and the target's state is not sufficient for such phenomenal character identity. For this reason, Coplan's affective matching condition is still too weak for empathy to be a source of phenomenal knowledge of the target's state. Consider the following example:

Stacie feels deep sadness (because a close relative died). Robert tries to empathise with her. He tries to simulate Stacie's mental state. The affective state he experiences during the simulation process is sadness, but only a mild feeling of sadness (like the sadness he experienced when he finished a book he enjoyed a lot).

---

6 Simulation is an imaginative process and thus faces a general problem with imagination as a source of knowledge: it might be that an imaginer forms beliefs as result of her imagination or simulation. It is, however, far from clear whether imagination or simulation is, or can be, a context of justification. In other words, if an empathiser simulates the target's state, it can of course happen that this simulation mirrors exactly the target's state. It is however unclear how the empathiser, on the basis of a mental simulation process, might be justified in believing that the simulated state is type-identical with the target's state. In general, in debates about empathy philosophers tend to assume that it is possible to gain knowledge by means of empathising with a target person. Here l will ignore the problem of justification and discuss only the question of whether empathy is a source of true phenomenal beliefs.

As we can now see, affective empathy comes in degrees. Robert's empathic state is phenomenally not exactly like Stacie's emotional state, but it might be closer than the empathic state of an empathiser who simulates an even milder state of sadness.

In the example, the condition of affective matching is met, because the difference between Stacie's state of sadness and Robert's empathic state is only a matter of degree. However, intuitively we doubt that Robert really knows what it is like for Stacie to feel her grief on the basis of his empathic engagement. Because Robert and Stacie's states differ phenomenologically in this example, Robert does not gain phenomenal knowledge of Stacie's state by means of experiencing his empathic emotion.[7]

7. Affective Empathy and Phenomenal Knowledge of Concurrent Emotions

Concurrent emotions are problematic for the optimistic view according to which we can learn what it is like to be in a target's state. An empathiser can try to gain knowledge of only one of the target's atomic states, or of the target's complex state. In the following, I argue that both enterprises face problems which result from concurrent emotions and the fact that they mutually influence each other's phenomenal character.

First, if it is correct that concurrent emotions mutually influence one another's phenomenal character, it is at least extremely likely that the phenomenal qualities of two people's type- and even content-identical emotional states will differ because of their different concurrent emotional states. Thus, an empathiser who simulates only one of the target's atomic (emotional) states is unlikely to acquire phenomenal knowledge of this atomic state, because of the possibility that the empathic emotion will have a different phenomenal character from the target's emotion. Hence, it is at the very least highly improbable that the simulated atomic emotional state of one person can be a source of phenomenal knowledge of the other person's atomic emotional state.

Second, for interpersonal understanding, it seems that, *ceteris paribus*, phenomenal knowledge of a complex conscious state is epistemically more relevant than phenomenal knowledge of a state which is only a part or constituent of this overall state. If a target person feels simultaneously sad and relieved, then phenomenal knowledge of the experience of simultaneously felt sadness and relief is epistemically better than phenomenal knowledge of either individual experience. Even if an empathiser who simulates each emotional state separately can have phenomenal knowledge of both token experiences, this empathiser would not necessarily have phenomenal knowledge of the state of a target who is simultaneously in both states, as the phenomenal character of the complex conscious state is more than its experiential parts.

Third, a simulation of a target person's complex conscious state is obviously more demanding than a simulation of an atomic affective state. I don't think that there are metaphysical reasons why a skilled empathiser could not simulate a target's complex state. However, the chances of achieving an exact simulation of the target's complex state with all its conscious components are extremely low. As a result, the simulated state will have fewer or different components than the target's complex state. Because the phenomenal character is a composition of its conscious components, it is then very likely that the phenomenal character of the empathiser's simulated state will differ from the phenomenal character of the target's complex

---

7   For the sake of simplicity, I ignore questions concerning sufficient conditions for gaining phenomenal knowledge of a target's mental state.

state. Hence, it is highly improbable that empathisers gain phenomenal knowledge of the target person's complex state.

There is a further problem: if simultaneous conscious states have an influence on each other's phenomenal character, as I have assumed, then the empathiser's awareness that she is simulating another person's state will also have an influence on her overall state as well. If this assumption is true, the phenomenal character of simulated complex states is also influenced by the empathiser's awareness that she is simulating. The phenomenal character of the empathiser's state in all cases involving awareness of the simulation is then phenomenally different from the target's state. It is also quite plausible to assume that the empathiser's awareness that she is simulating has an influence on the phenomenal character of each of the complex state's components. If this is true, even the simulation of an atomic state is phenomenally different to the target's atomic state. Thus, the simulation of either an atomic state or a complex state is not a source of phenomenal knowledge of the target's atomic or complex state.

8. Approaching Phenomenal Understanding: Propositional Knowledge and Explanatory Understanding

The empathy sceptic's claim that the epistemic goal of affective empathy is difficult, and sometimes even impossible to attain, is supported by the problem of concurrent emotions: if it is indeed massively unlikely that we can have phenomenal knowledge of another person's complex conscious state, this raises the question of whether phenomenal knowledge can play a decisive role, or indeed any role at all, in interpersonal understanding. It also raises the question of whether affective empathy is epistemically relevant.

An account of interpersonal understanding that answers these questions affirmatively has two options. One involves rethinking the standards of phenomenal knowledge. I have been assuming that we gain phenomenal knowledge of some mental state only by experiencing exactly this state. One could challenge this claim on the basis of the intuition that it seems implausible that a person who has experienced fear is likely to have no phenomenal knowledge at all of another's fear, merely because the two states are likely to have at least minimally differing phenomenal characters.

In this paper I will, however, adopt a different strategy: I will uphold the demanding standards I have set for phenomenal knowledge. For this reason I accept the sceptic's premise that affective empathy is indeed very often not a source of phenomenal knowledge. In contrast to the empathy sceptic, I will, however, claim that the epistemic goal of affective empathy is not phenomenal knowledge, but understanding. The idea is that phenomenal beliefs can have epistemic value even where they don't count as knowledge. This claim is grounded in a distinction between understanding and knowledge. In particular, it draws on arguments according to which (explanatory) understanding, in contrast to propositional knowledge, is not factive. The aim of the following section is to argue, in line with this non-factivist account of understanding, that there can be epistemic progress based on false phenomenal beliefs when these beliefs are relevantly close to a true belief. I call this epistemic progress "phenomenal understanding," and in the following take first steps toward an account of this variant of epistemic progress.

There have been intense philosophical debates on the relation between understanding and knowledge, debates which typically focus on explanatory understanding and propositional knowledge. While some authors argue

that understanding is a kind of knowledge (Kitcher 2002, Lipton 2004, Grimm 2006; 2012, and Khalifa 2012), others argue for the opposite claim (Zagzebski 2001, Kvanvig 2003, Pritchard 2009; 2010, Elgin 2004; 2017, Hills 2016). In these debates, understanding is typically compared with what can be taken to be the standard conception of propositional knowledge, according to which knowledge is (1) factive, (2) inconsistent with different types of luck, (3) cannot be formed on the basis of defeated evidence, and (4) can be passed on via testimony (see e.g. Hills 2016, 662).

Philosophers who argue that understanding is to be distinguished from knowledge do so on different grounds. In the following, I will focus on the fact that understanding, unlike knowledge, comes in degrees: that we can understand something better or worse, in more or less detail. This fact is arguably explained by two further features of understanding.

Firstly, many philosophers argue that understanding requires what Alison Hills calls cognitive control relative to the relevant set of beliefs about the object of understanding, i.e., that an understander possesses a set of cognitive abilities, such as drawing conclusions and giving or following explanations (Hills 2016). Degrees of understanding can be explained with reference to these abilities. That someone understands something better can mean that she is better in exercising one of the relevant cognitive abilities.

A second explanation of understanding's gradability refers to its belief component. Non-factivists claim that one's set of beliefs about the object of understanding do not need to be true in order to contribute to, or constitute, understanding. This is because the acquisition of false beliefs, they claim, can constitute cognitive progress. Elgin (2007) compares the cognitive status (1) of a person who believes that humans descended from great apes with (2) another person who believes that humans descended from butterflies, and (3) a person who believes that humans and great apes have a common ancestor who was, strictly speaking, not an ape. According to the state of the art, only the latter person has a true belief. However, Elgin argues that we should, and indeed do, distinguish the cognitive status of the first and the second person. The first person's belief is closer to the truth, and it would be fairly strange to say that this person does not understand anything about evolution at all. Hence, we should not only divide beliefs according to their truth and falsity, but also consider that a false belief can be closer or further away from the truth. If we accept that understanding can, and—according to Elgin—often does, involve false beliefs, degrees of understanding can be explained with reference to the degree of falsehood of its belief component. *Ceteris paribus*, an understander with a belief that is closer to the truth of a matter understands the matter better than an understander with a belief that is further away from the truth.

False beliefs not only play a role in individual understanding, but also in the sciences. In science, it is very likely that many of the current best theories will turn out to be false, even if the pessimistic view on scientific progress—according to which all of the current best theories may themselves be false—should be rejected. It is, however, consensus that some cognitive progress is being made in the sciences. By accepting that there can be cognitive progress on the basis of false beliefs, the non-factivity account of understanding allows that the sciences are able to attain a significant degree of understanding of their objects, and hence enables us to explain why talk of scientific progress is indeed correct (Elgin 2007; De Regt 2015).

Now, if understanding can be based on false beliefs, there is a certain danger that understanding will turn out to be a second-class cognitive state in comparison to knowledge. In particular, non-factivists need to explain understanding's epistemic value. This value is plausibly a result of the ability component of understanding:

although understanding is less demanding than knowledge, because the relevant beliefs are not subject to a truth requirement, it is more demanding, because it requires a set of cognitive abilities that operate on the relevant set of beliefs.

9. Non-Factivity: Phenomenal Understanding and its Belief Component

Debates on understanding and its relation to knowledge have focussed primarily on understanding why—i.e., explanatory understanding—and propositional knowledge. Clearly, though, if explanatory understanding can indeed be distinguished from propositional knowledge, for instance along the above lines, it certainly does not follow straightforwardly that all types of understanding and knowledge are distinct cognitive states. Hence, we need to have a closer look at the relation between phenomenal knowledge and what I am calling phenomenal understanding.

In analogy to explanatory understanding, I assume that phenomenal understanding has both a belief and an ability component. I will argue that phenomenal knowledge and phenomenal understanding are indeed distinguished by the latter's gradability, which, as in the case of explanatory understanding, results in turn from divergences relating both to the abilities involved and to the question of the states' factivity.

To begin with, it seems obvious that, like affective empathy, understanding another person comes in degrees, i.e., we can understand another person better or worse. Plausibly, this is not only true for explanatory understanding of another person; it is also true for interpersonal phenomenal understanding: it seems that even a false phenomenal belief can constitute cognitive progress, if it is close enough to a true phenomenal belief about the target's mental state. The basic idea is that an empathiser with a phenomenal belief about the target's state, which is relevantly close to a true phenomenal belief, understands the target, *ceteris paribus*, better than an understander with a non-phenomenal belief about the target's state. Take Peter Goldie's Irene, the icy-cold scientist, again. If we take the results of the discussion in section 7 of this paper into account, we have to assume that Irene is likely to form a false phenomenal belief about a target person's state. Nevertheless, she is in a better epistemic position to understand a target person who is afraid than she was before she had her first experience of fear.

Clearly, not every phenomenal belief about the target's state can bring epistemic progress. Think of the following (silly but coherent) comparison: understander A has only a non-phenomenal belief about the target's state. He recognises that the target is scared to death. The empathiser B has a phenomenal belief about the target's state. She also recognises that the target is scared to death and "connects" her phenomenal concept of joy. Obviously, understander B has a phenomenal belief about the target's state, and obviously it is a false phenomenal belief. With this false phenomenal belief, she is certainly not in an epistemically better position than understander A. There can be only epistemic progress on the basis of a false phenomenal belief if this belief is in a relevant sense close to a true phenomenal belief. What it means that a false phenomenal belief is in a relevant sense close to a true belief needs to be spelled out in more detail.

We gain phenomenal concepts by experiencing mental states with specific phenomenal qualities. Phenomenal qualities, or the phenomenal character of a state, can be more or less like the phenomenal qualities or character of another state. We can then assume, in an initial approximation, that the closer the phenomenal character

of the empathiser's state is to that of the target's state, the closer the empathiser's phenomenal concept is to the target's state. Hence, the empathiser's phenomenal belief about the target's state with the respective phenomenal concept is *ceteris paribus* closer to a true phenomenal belief about the target's state the more the understander's experience is (or was) phenomenally like the target's experience.

When concurrent states colour the phenomenal character of an emotional state, this introduces further parameters whose calibration could bring the phenomenal character of an atomic state closer to that of the target's: both the number of accompanying states and the proximity of each of their phenomenal qualities to the corresponding state of the target person will be relevant to the question of how close the empathic state is to the target's state. An empathiser who can simulate a significant number of the components of the target's complex state in a way that approximates each of their phenomenal characters will be able to form phenomenal beliefs which are relevantly close to true phenomenal beliefs. They will be in an epistemically better position than another person without such phenomenal beliefs.

10. Phenomenal Understanding and its Ability Components

In the previous section, I mentioned that differences in, or degrees of, explanatory understanding may derive from understanding's ability component. According to Hills, in order to understand a subject matter, an understander not only needs to have knowledge of it, but also needs what she calls "cognitive control." This means that the understander needs to have a set of cognitive skills or abilities which enable her to grasp the explanatory relationships between the propositions that she knows about the subject matter (Hills 2016). She presents a list of these cognitive skills intrinsic to explanatory understanding. On this list are abilities such as being able to follow and give an explanation, drawing conclusions, and so on.

Understanding's gradability derives, then, from the fact that an understander can be better or worse in exercising these cognitive abilities. In analogy to explanatory understanding, phenomenal understanding has components of cognitive skills as well. It is certainly debatable whether all items on the list are necessary, or whether some are too challenging. Hence, for an account of phenomenal understanding of another person, it might turn out that we don't need all of these skills for sufficient understanding. I assume, however, that at least some of the mentioned skills are also necessary for phenomenal understanding.[8]

For successful phenomenal understanding, the understander has to understand why the target person feels the way she feels. Here we can see clearly why phenomenal understanding is in this respect more demanding than mere phenomenal knowledge of another person's state, because phenomenal knowledge does not require cognitive control.

In order to phenomenally understand a target, the empathiser has a phenomenal belief about the target state *and* can, for example, explain why the target is in her state. In order to give such an explanation, the

---

8   Nick Wiltsher presents an account of what he calls "understanding what it is like to be (dis)privileged" (Wiltsher 2021). He develops an account of understanding of what it is like to be (dis)privileged and argues that the phenomenal knowledge of the experiences of (dis)privilege have to be accompanied by the cognitive skills we listed from Hills. He reformulates the list of skills in order to explain how a disprivileged person's experience is explained by social structures of disprivilege (Wiltsher 2021, 336–337). It is important, however, to note, that Wiltsher holds an optimistic view of empathy as a source of phenomenal knowledge (Wiltsher 2021, 323). His account of understanding-what-it-is-like is accordingly a factivist account of understanding. In the following, I will generalize Wiltsher's idea and apply it to interpersonal understanding in general in a non-factivist way.

empathiser needs further information about the target and her situation. Moreover, the empathiser needs at least basic folk psychological knowledge and needs to relate this knowledge to their phenomenal beliefs about the target's state.

Think again of the example taken from the novel *Middlesex*. An empathiser who wants to phenomenally understand Desdemona and her mixed emotions when she sees her husband lying dead on the kitchen floor gains phenomenal beliefs about Desdemona's state by means of affective empathy. From the description of the scene, and given some basic folk psychological knowledge, it is relatively easy to see why Desdemona experiences panic and grief in this moment. It is, however, difficult to understand why she feels happiness and relief. Further knowledge about her personality and her past can help. The author tells us that the death of her husband is the worst thing that could have happened to her, and that she was in deep worry that this could happen all through her adult life. With this additional information about Desdemona, we can understand her reactions: when she realises that this worst thing has now happened, she feels relief, because she believes that there is nothing to worry about in her life any more. However, the simulation process which takes place in affective empathy is not the source of knowledge or beliefs about these relevant additional facts of the target's situation or her biography. Thus, phenomenal understanding of another person cannot rely on affective empathy alone.

11. Conclusion

In this article I identified concurrent emotions as the main problem for the thesis that affective empathy can be a source of phenomenal knowledge of a target person's state. In the philosophical debate about empathy, the target person is often considered to be only in a single emotional state. I argued that this is a simplification, because we are often in a state of concurrent emotions. I distinguished atomic conscious states and complex conscious states, and assumed that emotions are atomic states in the explicated sense. One problem arises from the assumption that concurrent emotions have a mutual influence on their phenomenal character. An empathiser who simulates a single emotion without a view to its further attitudinal environment, when the target is in a state of concurrent emotions, will simulate a state with a different phenomenal character than the target's atomic emotional state. Further, I assumed that the set of conscious states we are in at a time has a phenomenal character, too. Phenomenal knowledge of the character of this complex state is of greater epistemic value in an attempt to understand another person than phenomenal knowledge of only a component of a complex state. It is because of the complexity that it is difficult, perhaps in many cases practically impossible, to simulate a target person's complex conscious state with the exact phenomenal character of the target's complex state. For this reason, affective empathy is, at least in many cases, not a source of phenomenal knowledge of the target's complex state either. I specified phenomenal knowledge as involving a propositionally structured true belief, where the propositional content contains a phenomenal concept. I argued that empathisers can gain phenomenal beliefs by empathising with a target. When these phenomenal beliefs are close enough to true beliefs, there can be cognitive progress. Finally, I claimed that false phenomenal beliefs can enhance one's understanding of the target person, if this belief is relevantly close to a true belief. Hence, phenomenal understanding is, in contrast to phenomenal knowledge, not factive. If, as I assume, phenomenal, like explanatory, understanding also has an ability component, an understander's capacity to explain why the target is in her atomic or complex conscious state is a candidate for such a component. Further research needs to examine which abilities are involved in interpersonal phenomenal understanding and how exactly they contribute to its particular depth.

References

Bailey, O. 2022. "Empathy and the Value of Humane Understanding." *Philosophy and Phenomenological Research* 104 (1): 50–65.

Balog, K. 1999. "Conceivability, possibility, and the mind-body problem." *Philosophical Review* 108: 497–528.

———. 2009. "Phenomenal Concepts." In *The Oxford Handbook of Philosophy of Mind*, edited by B. McLaughlin, A. Beckermann, and S. Walter, 292–312. Oxford University Press.

Batson, C. D. 2009. "These Things Called Empathy: Eight Related but Distinct Phenomena." In *The Social Neuroscience of Empathy*, edited by J. Decety and W. Ickes, 3–16. MIT Press.

Bayne, T. J. and D. J. Chalmers. 2003. "What is the Unity of Consciousness?" In *The Unity of Consciousness. Binding, Integration, and Dissociation*, edited by A. Cleeremans, 23–25. Oxford University Press.

Boisserie-Lacroix, C. and M. Inchingolo. 2021. "Empathy for a Reason? From Understanding Agency to Phenomenal Insight." *Synthese* 198: 7097–118.

Coplan, A. 2011. "Understanding Empathy: Its Features and Effects." In *Empathy: Philosophical and Psychological Perspectives*, edited by A. Coplan and P. Goldie, 3–19. Oxford University Press.

De Regt, H. W. 2015. "Scientific Understanding: Truth or Dare?" *Synthese* 192: 3781–97.

Döring, S. A. 2007. "Seeing What to Do: Affective Perception and Rational Motivation." *Dialectica* 61(3): 363–94.

Elgin, C. Z. 2004. "True Enough." *Philosophical Issues* 14: 113–21.

———. 2007. "Understanding and the Facts." *Philosophical Studies* 132: 33–42.

———. 2017. *True Enough.* MIT Press.

Eugenides, J. 2002. *Middlesex.* Farrar, Straus, and Giroux.

Goldie, P. 2002. "Emotions, Feelings and Intentionality." *Phenomenology and Cognitive Sciences* 1: 235–54.

Green, O. H. 1992. "*The Emotions: Philosophical Theory.* Kluwer Academic Publishers.

Grice, H. P. 1941. "Personal Identity." *Mind* 50: 330–50.

Grimm, S. R. 2006. "Is Understanding a Species of Knowledge?" *British Journal of Philosophy and Science* 57: 515–35.

———. 2012. "The Value of Understanding." *Philosophy Compass* 7: 103–17.

Heavey C. L., N. L. Lefforge, L. Lapping-Carr, and R. T. Hurlburt. 2017. "Mixed Emotions: Toward a Phenomenology of Blended and Multiple Feelings." *Emotion Review* 9 (2): 105–10.

Hills, A. 2016. "Understanding Why." *Noûs* 49: 661–88.

Hoffmann, M. L. 2000. *Empathy and Moral Development: Implications for Caring and Justice.* Cambridge University Press.

———. 2011. "Empathy, Justice, and the Law." In *Empathy: Philosophical and Psychological Perspectives*, edited by A. Coplan and P. Goldie, 230–54. Oxford University Press.

Jackson, F. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127–36.

Kaupinnen, A. 2017. "Empathy, Emotion Regulation, and Moral Judgment." In *The Routledge Handbook of Philosophy of Empathy*, edited by H. Maibom, 2015–27. Routledge.

Khalifa, K. 2012. "Inaugurating Understanding or Repackaging Explanation?" *Philosophy of Science* 79: 15–37.

Kind, A. 2020. "What Imagination Teaches." In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, edited by J. Schwenkler and E. Lambert, 133–46. Oxford University Press.

Kitcher, P. 2002. "Scientific Knowledge." In *The Oxford Handbook of Epistemology*, edited by P. K. Moser, 385–408. Oxford University Press.

Kvanvig, J. 2003. *The Value of Knowledge and the Pursuit of Understanding.* Cambridge University Press.

Larsen, J. T. 2017. "Introduction to the Special Section on Mixed Emotions." *Emotion Review* 9 (2): 97–98.

Lipton, P. 2004. *Inference to the Best Explanation.* Routledge.

Maibom, H. 2017. "Affective Empathy." In *The Routledge Handbook of Philosophy of Empathy*, edited by H. Maibom, 22–33. Routledge.

Marks, J. 1982. "A Theory of Emotion." *Philosophical Studies* 42 (1): 227–42.

Melnyk, A. 2002. "Papineau on the Intuition of Distinctness." *SWIF Philosophy of Mind Review December 2002.*

Nida-Rümelin, M. 1998. "On Belief about Experiences. An Epistemological Distinction Applied to the Knowledge Argument against Physicalism." *Philosophy and Phenomenological Research* 58 (1): 51–73.

Papineau, D. 1993. *Philosophical Naturalism.* Blackwell.

———. 2002. *Thinking About Consciousness.* Oxford University Press.

———. 2007. "Phenomenal and Perceptual Concepts." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by T. Alter and S. Walter, 111–44. Oxford University Press.

Prinz, J. 2004. *Gut Reactions: A Perceptual Theory of Emotion.* Oxford University Press.

Pritchard, D. 2009. "Knowledge, Understanding, and Epistemic Value." In *Epistemology*, edited by A. O'Hear, 19–43. Cambridge University Press.

———. 2010. "The Value of Knowledge: Understanding." In *Knowledge and Understanding*, edited by D. Pritchard, A. Millar, and A. Haddock, 3–90. Cambridge University Press.

Roberts, R. C. 2003. *Emotions: An Essay in Aid of Moral Psychology*. Cambridge University Press.

Roseman, I. 2017. "Transformative Events: Appraisal Bases of Passion and Mixed Emotions." *Emotion Review* 9 (2): 133–39.

Searle, J. R. 1983. *Intentionality: An Essay in the Philosophy of Mind.* Cambridge University Press.

Smith, J. 2017. "What is Empathy For?" *Philosophical Studies* 194: 709–22.

Spaulding, S. 2017. "Cognitive Empathy." In *The Routledge Handbook of Philosophy of Empathy*, edited by H. Maibom, 13–22. Routledge.

Stueber, K. Forthcoming. "The Structure of Rational Agency and the Phenomenal Dimension of Empathic Understanding." In *Empathy's Role in Understanding Persons, Literature, and Art*, edited by T. Petraschka and C. Werner, 119–37. Routledge.

Sundström, P. 2011. "Phenomenal Concepts." *Philosophy Compass* 6 (4): 267–81.

Tappolet, C. 2016. *Emotions, Values, and Agency.* Oxford University Press.

Wiltsher, N. 2021. "Understanding What It's Like To Be (Dis)Privileged." *Pacific Philosophical Quarterly* 102 (2): 320–56.

Zagzebski, L. 2001. "Recovering Understanding." In *Knowledge, Truth, and Duty*, edited by M. Steup, 235–51. Oxford University Press.

Shame, Guilt, and the "Morality System"

Edward Harcourt - Keble College, Oxford

Abstract

Arguably the differences between guilt and shame are exaggerated in the literature, especially with respect to the relationship of each to morality. Some fresh examples of shame are presented which point in the same direction, but which also indicate a puzzling dualism within the structure of shame. Furthermore, via Williams's idea of the 'morality system', an explanation is offered of survivor guilt which does not compromise its distinctiveness (that is, without resorting to the hypothesis that the survivor has harmed someone by surviving).

1.

Enemies of the "morality system"[1] have tended to be friends of shame. This was certainly true of Wollheim[2]—though the phrase "the morality system" came later—and with some qualifications it is also true of Williams, who of course introduced the phrase. The phrase notwithstanding, it's hard to say exactly how the different bits of the morality system are supposed to form a *system*—possibly because they don't. But whether it is really a system or not, I shall think of this undoubtedly "distinctive formation within the ethical" as giving special significance at least to the ideas of moral obligation; of the internal or non-other-involving as opposed to the other-involving or social; and of autonomy, moral responsibility, and desert. Guilt and the morality system supposedly belong together because, on this way of thinking, guilt is the emotion that goes with one's

---

1   Williams 1985, esp. ch. 10. The qualification I have in mind is one of Williams' points: it's not that Greek ethics was for shame and against guilt, but rather that it didn't have the distinction. So their "shame" cannot *quite* have been the modern notion, even if the Greek word *aidos* overlaps with our uses of "shame." By forcing the distinction between shame and guilt, the morality system obscures the possibility of a psychological economy regulated by an emotion that is not clearly one or the other. For the morality system, see also Williams 1993, 5 ("freedom, autonomy, inner responsibility, moral obligation").

2   See Wollheim 1983.

own internal judgment that one is responsible for failing to meet a distinctively moral obligation. Shame, by contrast, is said to have a much wider range of intelligible objects, to be more loosely associated with responsibility, less sensitive to one's own judgment and more sensitive to the judgment of others.

I've argued in previous work that the contrasts along these various dimensions between guilt and shame are exaggerated and sometimes ill-grounded, and so do not underwrite a credible normative distinction between the two emotions, distinct though they surely are.[3] The point was not, however, to rehabilitate shame for the morality system but rather to argue that both guilt and shame have a proper place outside of it. In this paper I want to raise two challenges to that ambition, and to try to respond to them. The first challenge is that there is after all a distinctive *form* of shame which belongs especially closely with moral obligation and responsibility and which, while still social, is less thoroughly other-involving than other recognisable forms of the emotion. The second challenge is that there is a distinctive and puzzling form of guilt—survivor guilt—which, while it looks at first glance like more evidence against a tight connection between guilt and morality, is really only intelligible when seen in relation to the morality system.

2.

I'll begin, however, by pressing further the point that the differences between guilt and shame are overdone, by taking issue with Williams's view, in *Shame and Necessity*, that shame is connected especially closely—as he says guilt is not—with being seen. "Shame and its motivations always involve in some way or other an idea of the gaze of another" (1993, 82), he says, and again:

> The basic experience connected with shame is that of being seen, inappropriately, by the wrong people, in the wrong condition. It is straightforwardly connected with nakedness, particularly in sexual connections. (1993, 78)[4]

But this seems mistaken, for two reasons. First of all, consider Wollheim's nice example of his own shame at *overhearing* himself being discussed (and his parents being overheard discussing him) (2004). The case shows at the very least that the connection is between shame and being *observed*, in a range of modalities—not just sight; or more loosely still, between shame and being known about. The newspapers publish some indiscretion of mine and I am ashamed, but they need not have learnt about the indiscretion perceptually, nor is my exposure by way of being perceived (because it's by way of publication).[5]

Secondly, to the extent that shame is connected with being seen, so is guilt. An obvious example is the secret adulterous affair, which gives rise to the evidenceless worry on the part of the guilt-ridden adulterer that he has been seen: is there some CCTV footage of him that he didn't think of? But perhaps, one might think, this connection between guilt and being seen is specific to sexual misdemeanours; sexuality takes us to nakedness, and nakedness takes us to shame. So the connection between guilt and being seen is owed to the fact that

---

3 See my "Moral Emotions, Autonomy and the 'Extended Mind'" (2016) and "Guilt, Shame, and the 'Psychology of Love'" (2007).

4 See also Maibom 2010, 569.

5 In such a situation, I may be inclined to *say*, "It's as if everything I do is open to public view." This suggests that if there is a special connection between shame and vision, it is because of the special aptness of vision to *symbolise* experiences associated with shame. Compare also our inclination to say we feel as if parts of ourselves have been seen which are not in fact (ordinarily) visible ("My bones are not hid from thee"; Psalm 139:14), though the strength of the link between these locutions and shame is a matter for further discussion. Thanks to James Laing for prompting these thoughts.

shame is implicated somehow after all. But the association between guilt and being seen does not rely on sexual examples. I am due to leave Venice in a couple of hours and my 48-hour Vaporetto pass expires. But I have one more trip to make, in order to check out of my hotel. So I take that last trip without a ticket. My subsequent torment is not driven by the thought of the harms I have caused by what I have done: I certainly *don't* think, "How unfair on the other passengers," or "Will the Vaporetto company go bust?" Instead, I imagine one person after another getting onto the boat, falsely as it turned out, to be a ticket inspector—is that blue jacket some kind of uniform?—even though I have not once seen a ticket inspector in Venice. That is, my only thought is that I will be exposed for having committed an albeit trivial transgression. Indeed, it is almost as if, having committed the transgression, I am convinced I have *already* been exposed—that is, have already been seen by inspectors, though I am unable to identify them—hence the quasi-paranoid imaginings. But what I feel is guilt, not mere fear of detection, nor fear of shame upon detection.

I want to turn now, however, to the first of my two challenges. In outline, the challenge is that even if we absorb both of the above points, there isn't just *one* connection between shame and being seen (or observed, or known about), so, if the connection between shame and these things is meant to affirm a connection between shame and the social, no *one* connection between shame and the social is thereby affirmed. Indeed, the connections are so different that we may end up wondering whether we have a single emotion here. Because one of the two connections I shall distinguish looks set, *pace* Wollheim and Williams, to rehabilitate the connection between shame and the morality system, this is a challenge to the idea that shame points to a psychological economy which is not dominated by morality.

3.

The challenge is best pursued by way of examples. Example 1: A student has inadvertently left a curtain open and, because my room overlooks hers, I accidentally catch sight of her undressing and she sees that I have seen. She is ashamed. Example 2: I catch a student covertly cheating in an exam, and again the student is ashamed. There is a surface similarity between examples 1 and 2: in each case someone feels shame upon being seen, and in each case they are ashamed at something which falls within the wide range of intelligible objects of shame (respectively, nakedness in an inappropriate context, and cheating). But further examination reveals gross differences of structure between the two cases.

In 1, while there is an expectation on the student to cover herself, there is an equally strong or stronger expectation on me to avert my gaze. Moreover, if indeed I do avert my gaze, part of what motivates me to do so is the shame *I* reasonably feel at seeing *her*. Thirdly, the expectation on the student to cover herself stems at least in part from the expectation that she should deliver me from my shame at seeing her, an expectation to which she is subject even if my catching sight of her isn't in any way her fault (in some very weak sense one could say it is her fault if she absent-mindedly forgot to close the curtain, but not if she always has the curtain open at a certain time of day, confident in the expectation that no one will see her, and then for some reason I arrive in my office unexpectedly early). Fourthly, if I—shamelessly—fail to avert my gaze, the student can retreat to the thought that there is nothing wrong with being naked: we were all born naked, all look the same under our clothes, and so on. This thought—though it may take some toughness of mind to think it—can

serve as an effective defence for her against her shame.[6] Finally, and relatedly, if she doesn't see that I have seen her—and assuming that we are not ladies' finishing school fanatics who think it is shameful just to *be* naked[7]—there is no occasion for her to feel shame. I'll come back to this final point later. For now, I simply note that if the student deploys the tough-minded defence, it does not lift the expectation on me to feel shame if I fail to avert my gaze (though if I'm shameless enough, I won't actually feel it).

By contrast, none of this structure of reciprocal expectations is present in case 2. It isn't shameful for me to observe the student cheating, even though her act is shameful. I might feel pained to observe her, for example, if she is a vulnerable or foolish student. But to feel pained on her behalf isn't to feel shame, and there is no expectation on one even to feel pained: one might reasonably feel anger or exasperation at the student instead. There is certainly no expectation on the student to protect me from shame by concealing her behaviour more effectively. Is there anything she can retreat to in order to defend herself against shame? She could protest that there's nothing wrong with cheating—but, other things being equal, that isn't true. Of course, nobody is saying that the student in 2 *actually will* feel shame—she might be shameless and feel nothing at all. All that's being claimed is that cheating is an appropriate object of the student's shame.

4.

What, then, are we to do with the contrasts between 1 and 2? One common response is that, while both kinds of case surely exist, this does not show the existence of two forms of shame with different structures. That's because, so the response goes, while case 2 is a case of shame, case 1—the case with the complicated structure of reciprocal expectations—is merely a case of embarrassment. Now, without for the moment commenting further on this response, it seems to me that there is a plausible, though not decisive, argument that the shame in case 1 is indeed what we may call embarrassment, and that this is not true of the shame in case 2. The test is to think through cases 1 and 2, and ask whether shame and embarrassment can be co-present but with distinct objects.

In case 2, it appears that shame as already described in that case and embarrassment can be co-present, at least in the following ways.[8] First, it is conceivable that the student might feel embarrassment (that it's her familiar teacher who has caught her, rather than an anonymous official) alongside shame as already described (that she has done something wrong). Secondly, one of the many emotions I might feel on detecting her cheating, besides anger, disappointment, and pain on her behalf is, of course, embarrassment ("I wish it hadn't been me to notice that she was cheating"). Call this imagined variation on the case 2a. So, in 2a, is the thing that runs alongside shame, as already described, the emotion (whatever it's called) that we discerned in case 1, or something else? Here are two apparent reasons for thinking it's something else. In case 1, it is the student's emotion that triggers my experience of the same emotion at observing her, whereas in 2a my embarrassment may be triggered whether she feels it or not. And in 2a there is no expectation that she should protect me from my embarrassment at observing her. But there are replies to both thoughts. If she really isn't embarrassed—because it's a matter of indifference to her that she's been caught by me in particular—my own embarrassment

---

6   It's worth saying, parenthetically, that this structure of mutual expectations is changed significantly if I see the student because I somehow deliberately intrude upon her, and also that sex difference may play a role here. But the fact that one can upset the example by varying the details is irrelevant—I only need one example of shame with this structure to make my point.

7   Compare with the nuns who keep their clothes on in the bath (Russell 1927). See also Williams 1993, 82, n. 20.

8   Thanks to Tom Sinclair for pressing me on the implications of this and related variations on the case.

ought to lapse, just as my emotion might lapse in case 1 if I discover that the naked student is in fact taking part in some convoluted art piece in which people are intended to see her, in a context in which this would normally happen by accident. Secondly, while in 2a it's not in the student's power to prevent me from being embarrassed, it would make sense for her to *apologise* to me for having caused me embarrassment, a ground for apology that floats free of whatever apology she might make for cheating. The fact that apology would make sense in this context could be argued precisely to reflect the (unfulfillable) expectation on her to protect me from embarrassment. Can we do something similar for case 1, that is, reimagine case 1 with the co-presence of both shame—that is, shame as already described in *that* case, the thing with the complicated structure of reciprocal expectations—*and* embarrassment? But what *else* is there in case 1 aside from the student's being seen naked, my seeing the student naked, etc., to be embarrassed about? Surely the imaginative exercise is impossible.

These replies, to the extent that they are effective, make it look as if what in 2a we're calling embarrassment is what we called shame when describing case 1. It's important to recognise, however, how little is accomplished by this classificatory move. The key question is whether embarrassment is or is not a sub-variety of shame. If it is, then if what we began by calling shame in case 1 *is* embarrassment, the differences between cases 1 and 2 confirm the existence of two forms of shame with different structures, rather than providing a way to avoid that conclusion. But the classificatory move leaves that key question open, for the fact that cases 1 and 2 attract different verbal labels, if indeed they do, is very weak evidence that they exemplify different emotions (compare "hate" and "loathing"). Thus invoking the idea of embarrassment provides no refuge from the puzzle that shame seemingly can have two distinct structures. I shall therefore pursue the challenge presented by the contrast between cases 1 and 2, leaving the notion of embarrassment to one side.

One might conclude that the real difference between 1 and 2 is this: what's shameful in 2 is not *being seen cheating* but simply cheating. By contrast, what's shameful in 1 isn't being naked, but *being seen naked*. In 1, the object of shame is two-place: the person of the observer is partly constitutive of the object of the emotion. In 2, it's one-place.[9] We will have occasion to revisit this provisional conclusion below.

5.

One reason the contrasts are challenging is the following: in cases like 2 the observed act is shameful—that is, deserving of shame—whether or not the person observed actually feels shame (and they often don't). And of course something like this also works in the opposite direction: many people will feel shame in a case like 2—even if they are not actually observed. So we need to ask whether there is anything left in cases like 2 of the supposed connection between shame and being seen, even once this is commuted into a connection with being perceived (in whatsoever modality) or simply being known. Is the difference between 1 and 2 really a difference in the *kind* of connection between shame and knowing/being known, or is it in fact *only* in cases like 1 that any such connection obtains? That difficulty matters, because it was supposedly via the link to sight (etc.) that the social or other-regarding character of the emotion was established.

---

9 The difference between 1 and 2 survives if we swap out vision for some other sense-modality or none. In Wollheim's example, an audience is seemingly built in to the object of shame, and built in because there is an implicit audience restriction built in to the conversation: it's not that there is anything *per se* shameful in parents discussing their child, but their discussion is not "for his ears," and so being overheard—and overhearing—is shameful. So Wollheim's example has the same structure as 1, though it's about hearing and not seeing. Similarly, in a parallel variation on 2, I might catch a student cheating by overhearing something or reading an email exchange.

To this let me add a second and more important reason. There's an argument to be made that cases like 1 are somehow dependent—"merely dependent," perhaps—on the opinion of others. That's why they can be magicked out of existence by the tough-minded defence that there's nothing wrong with being naked. The same argument says that cases like 2, by contrast, are about wrongs that are independent of the opinion of others, indeed dependent on norms that the opinion of others ought to track (and is at fault if it doesn't). So, the thought continues, what really unifies cases of the kind of shame that can't be magicked away by the tough-minded defence is that they are cases of *moral* shame. A. W. H. Adkins, who in Williams's writing exemplifies allegiance to the "morality system" (1993, 81), declared his opposition to shame on the grounds that it is *merely* social, and so superficial, in contrast to the properly moral emotion of guilt. That is surely a mistake. But one effect of distinguishing cases 1 and 2 might be that we carve off one class of cases—let's call these type 1—as thoroughly social (whether or not that makes them of lesser significance), but in so doing rescue a second class of cases—call these type 2—that meet all of Adkins's standards of supposed moral maturity. That's a more troubling reason, because a central part of the defence of shame as an emotion characteristic of a mature practical consciousness, and the attack on the morality system that follows, was precisely the breadth of its proper objects. Distinctively moral objects, on that account, have no special status, whereas on the account now being envisaged, for type 2 cases they do.

So, to deal with the first problem first: what, if anything, is left of the connection between shame and being seen (or observed, or known) in type 2 cases? In brief, my answer is: nothing, but the social character of shame doesn't depend on it.

In type 2 cases, shame is a social emotion because it's essentially other-involving in the following sense:[10] where it is regulating a well-functioning practical consciousness, its operation depends on a community of values between the subject of the emotion and others around them.[11] So I will feel ashamed of something in your presence only if you and I share values in the light of which it's shameful. If you despise cheating and I don't, your presence won't make any difference to whether I experience shame or not. But if we are on the same page about these things, your presence will make a big difference—amplifying the emotion, even if it would be triggered without you there. Since shame helps to regulate our behaviour, the disposition to feel shame is thus one example of the way in which we rely in self-regulation on other people.

What about the fact that I can feel type 2 shame on my own? The best analogy for this phenomenon is the division of linguistic labour (Putnam 1973, 699–711): we use technical words in ordinary conversation (for body parts, chemicals, meteorological phenomena, and so on), conveying knowledge thereby, but also without opening a dictionary or consulting an expert. But still, the practice of using such words—whose full explanation lies beyond most speakers' ability—depends on there being dictionaries and experts we can track back to if we get stuck. The experts are implicit in the practice of using such words—the practice depends on

---

10 See Williams 1993, 83: "The bonding, interactive effects of shame" connect with the fact that "an agent will be motivated by prospective shame in the face of people who would be angered by conduct that, in turn they would avoid for the same reason." For other sources that make the same point, see Maibom 2010, 570. And see also my 2016. I also argue that the same can be said of guilt, but never mind that here.

11 The qualification about a well-functioning practical consciousness is important, because there are also cases of shame where my liability to the emotion does not depend on values in common—e.g. where, attending a family occasion, I am made to feel ashamed by my failure to conform to family norms which I have long since sincerely rejected. There are several different explanations of cases like this: habit takes a while to catch up with practical reason; my detachment from the family is only partial because while I reject some values of theirs I also still want to be accepted by them, and so on. But I take it these are all in some sense cases where practical consciousness is *not* functioning well. Thanks to Charlie Kurth for pressing this point.

them—even if they are not *actually* consulted, in every (or indeed most) exchanges. Something similar, I would argue, goes for shame: it doesn't detract from the social character of the emotion that the relevant audience for shame—i.e., those with whom we hold values in common—need not actually be present when the emotion is experienced.[12]

This observation works as a defence of the claim that shame is essentially social despite the fact that it can be experienced alone. However, the idea of the sociality of shame in play here seems very different from the idea that shame is specially connected with being seen or indeed with being observed. Indeed, the latter idea may in type 2 cases just be an obscure way of registering an important but importantly different fact about shame, namely its sociality or other-involvingness in the sense I've explained.

6.

I'll move on to the second problem—whether the difference between type 1 and type 2 shame rehabilitates, for a sub-class of cases, the link between shame and the morality system—in the course of this section. But, to introduce it, I want to pause first on Williams in *Shame and Necessity*, and in particular on his discussion of Sophocles' *Ajax*.

Williams is, in my view, quite right when he insists on the "bonding, interactive effects of shame," which he connects with the fact that "an agent will be motivated by prospective shame in the face of people who would be angered by conduct that, in turn, they would avoid for the same reason" (1993, 83).[13] That's just the point already made that the successful operation of shame depends on a community of values between the subject of the emotion and others around them. But, as if continuing exactly the same point, Williams then turns to Ajax. Ajax, it will be recalled, kills himself, having made himself absurd slaughtering a flock of sheep under the illusion that he was avenging a slight inflicted on him by his fellow warriors (Sophocles 2008). Ajax laments:

> What countenance can I show my father Telamon?
> How will he bear the sight of me
> If I come before him naked, without any glory?

Williams rightly points to "the interlocked expectations between [Ajax] and the world," but as if they exemplified no more than the community of values between Ajax and his warrior peers (most saliently, his father) that we've already mentioned. That community of values is already present, of course, in case 2. But Ajax's situation is ostensibly much more complicated than that. His concern is not just that he will rightly feel shame before his father (which depends on their holding values in common), but that his father will feel shame to see him in a shameful condition ("How will he bear the sight of me?"), and that it's a requirement upon him—which he fulfils by his suicide—to protect his father from the shame of seeing him. That looks

---

12 In my 2016 "Moral Emotion, Autonomy and the Extended Mind," I argued for the further point that there's no inference from the thought that shame is social, or other-involving, to the conclusion that it is heteronomous or otherwise immature: on the contrary, on a proper conception of autonomy, autonomy involves getting other-dependence right, not doing without it.

13 In fact I am not so sure about motivation by *prospective* shame—unless the point is very carefully expressed, it can look as if what's motivating is the prospective unpleasantness of shame, when what really motivates is the prospective shamefulness of the act, to which that shame would be a response. See my 2013, "Happenings outside one's Moral Self: Reflections on Utilitarianism and Moral Emotion," 239–258. However, I think this is merely an incautious formulation on Williams's part as the inside/outside distinction in "Happenings" is due to him.

very like the expectation on the student in case 1 to protect me from the shame of seeing her naked. How, in terms of our distinction between case 1 and case 2, is Ajax's case to be understood?

The fourth feature of case 1 is a good starting point, namely the availability of the "tough-minded defence" that there's nothing really wrong with being naked. Ajax was surely not capable of the thought that there is nothing really wrong with looking utterly foolish. But let's nonetheless explore the idea that Ajax's case too is a case of type 1 shame. An obstacle to this view is my earlier thought that the object of shame in 1 is not being naked, but being seen naked. For it is surely wrong to say Ajax is ashamed just of *being seen* to have done something foolish—though he is ashamed of that, he's ashamed of it because he is more basically ashamed of his folly. So, revisiting that earlier thought, should we by analogy after all say that the object of shame in 1 is— thanks, perhaps, to the context—nakedness itself?

In favour of this way of viewing things, it may help to observe that, like the female student's reactions in 1, Ajax's reactions are structured by a conception of a social role—in his case, that of the aristocratic warrior.[14] One reason for the seemingly readier availability of the tough-minded defence in the student's case is that, in our world, social roles are more easily taken up and vacated than in Sophocles', because we can make more sense of a distinction between the self and the various social roles we can occupy. So, while the student may very well structure her behaviour around what she sees as appropriate for a woman (or a young woman, a Western woman, etc.), she can also in a moment of crisis think "to hell with that role and its demands" in a way Ajax cannot. Conversely, one doesn't have to think very far back in our own history to find women for whom that defence might not have been possible, and for whom an episode of the kind I've envisaged might have been life-changing, if not life-ending, in a way Ajax's folly was for Ajax. If one did even in such times offer the defence, it would have been on pain of a kind of social obliteration, exiting forever a social role ("lady," perhaps) without having any clear alternative to it.[15] Similarly, even if Ajax couldn't think of himself outside the warrior role at that moment, one can imagine an alternative (and not much happier) outcome for an Ajax-like character, namely that he offers the defence—"anyone can suffer a delusion!"—and exits forever his warrior role to continue a life of role-less obscurity. Where roles are tighter and fewer in number than they are for us, exiting one makes one a "poor bare forked animal":[16] there is not the dignity of what we might call the "human role" to default to, which the student invokes, perhaps characteristic of more modern times.[17] On this account, the readier availability of the defence in the student's case but not in Ajax's does not mark a deep difference between them. The complex structure of reciprocal expectations in Ajax's case would thus speak in favour of classifying Ajax's as type 1 shame.

One lesson from Ajax's case, on that view, would be that the distinction between type 1 and type 2 cases does not mark a difference between the trivial and the important: the object of shame in type 1 cases *can* be something trivial, but it can also be as important and all-consuming as these things get.[18] Another lesson is that the availability of the tough-minded defence belongs with a society in which people are less tightly bound to social roles than they were in Sophocles' day or, otherwise put, that the student can, while Ajax cannot, retreat to the "human role" without bringing meaningful life to a close. That defence also brings the elaborate structure of shame in type 1 cases crashing to the ground. But now, to take up again the question whether the

---

14 For more on shame and social roles, see Deigh 1983, 224–45.
15 Perhaps Aubrey's story of Elizabeth I and the Earl of Oxford gives some idea: see Aubrey 1957.
16 Shakespeare 1608/1997, III.4.
17 Cf. Korsgaard 1996, 101: "You are a human being, a woman or a man, a member of a certain ethnic group," etc. It's at least the "human being" part that the student does, but Ajax does not, have the opportunity to default to.
18 That, incidentally, is another reason why the word "embarrassment" might not be a fit for type 1 cases, at least not for all of them.

difference between type 1 and type 2 shame rehabilitates the link between shame and the morality system, isn't the very concept of the moral coeval not just with the availability but with the occupancy of the "human role"—as of course saliently in post-Kantian thinking? So—the thought continues—while type 1 cases need not be trivial (and indeed can be hugely important), they *cannot* be moral, because the moral comes on the scene only once the elaborate structure of role-dependent expectations is, as in the student's self-defence, left behind. While this wouldn't yet show the worrisome conclusion—that the distinction between type 2 and type 1 cases aligns with the distinction between the moral and the non-moral, thus affirming a connection between shame and the morality system, albeit for a restricted class of cases—it would get us halfway there, because it would force us to isolate a form of essentially non-moral shame. All that's needed to complete the argument is that—as case 2 seems to be—the other side of the distinction is occupied solely by moral cases.

But notwithstanding its complex structure of reciprocal expectations, is Ajax's case really—like the naked student's—a case of type 1 shame? There is a certain strain in saying that the object of the student's shame in case 1 is nakedness rather than being seen naked: though there are no doubt finishing-school fanatics around, imagining her to be one is quite unnecessary to imagining the complex structure. Moreover, the structure of reciprocal expectations in Ajax's case can be explained in another way, via his and his father's mutual identifications. On that account, Ajax's case would not resemble case 1, but another variation on case 2—call it 2b—in which our student is caught cheating in the exam and cannot face her parents afterwards, partly because she is ashamed of what she has done but partly because she knows they will be ashamed of her. This should not surprise us: to the extent that the norm against cheating has a grip on people, there will be some who take pride in observing it and feel shame at failing to do so, and this pride and shame extends to those, such as their children, with whom they are closely identified.

Now, although in 2b the student's seeing her parents will make things far worse, she knows full well that they will be ashamed of her (and that she has failed them in making them ashamed of her) even if she doesn't see them. Moreover, as in the simple case 2, retreating to the "human role" is no defence because, as she knows and knows her parents know, cheating is wrong. It's tempting to conclude that what disarms the tough-minded defence in 2b is that the wrongness of cheating is role-independently—or, as one might say, morally—wrong.

If Ajax's case is relevantly like 2b rather than like 1, two conclusions follow. First, *pace* Williams, the relationship of shame to being seen or observed or simply known dwindles in that case almost to zero. The only thing sustaining the structure of reciprocal expectations in Ajax's case is the set of values held in common between the subject of shame and its audience—as in the simple case 2—plus identification. The similarity with case 1, where being seen (etc.) really *does* have a constitutive role, is an appearance only. More importantly for the worry about the morality system, just as case 1 looks as if it helps us to identify a form of essentially non-moral shame, case 2 makes it look as if we might be forced to identify an essentially moral form. But now we have a reply to that problem. No tough-minded defence is available to Ajax, but neither is it to the student in 2b (or 2). But for all that, insofar as we understand the distinction between moral and non-moral at all,[19] neither suffering a delusion nor making a fool of oneself are moral failings, any more than competitive failures (such as losing a race or a piano competition), suffering other indignities for which one bears no responsibility, or standing features of oneself for which, again, one is not responsible (a birthmark, a stammer, etc.).

---

19 Of course, there is a degree of discomfort for the critic of the morality system to assume, for the purposes of argument, that the distinction between moral and non-moral is well understood (is type 2 shame distinctively *what?*). But it seems unavoidable at this stage in the dialectic.

In case the unavailability of the tough-minded defence is thought to be local to the ancient context, consider a child's poor performance in a race of a kind in which they normally do well. Some families nowadays may think of these occasions in an Ajax-like way: the poor performance has disgraced the family, the child knows this so not only feels ashamed but an obligation to protect the parents from the shame of witnessing (or otherwise finding out) about the performance, and so on. But even if the family doesn't think this way, as in the cheating case, it won't help to tell the child that athletics is unimportant and so it is *not really shameful*: either the child will detect the insincerity or, if it is sincere, will make things worse (that is, by discovering that their most important audience doesn't care how well they do). That's just the "community of values" point over again. If "it's not really shameful" expresses anything coherent at all, it is simply registering that the failure is not one's fault. But the fact that that is, while true, so unhelpful underscores the point that type 2 shame has intelligible non-moral objects. So type 2 shame is not an essentially moral form of the emotion.

In sum, our original puzzle—that there appear to be two forms of shame with different structures, as exemplified by cases 1 and 2—remains, and the sense in which shame is a social emotion is very different in cases of the two types. The link between shame and seeing/being seen is weaker in either type 1 or type 2 cases than is often said. But while in type 1 cases some connection to others' observation or knowledge is essential to characterising the object of shame, in type 2 cases the alleged connection to being seen (etc.) dwindles almost to nothing. However, while it is arguable that type 1 cases can't be moral, type 2 cases—the first challenge presented by this structural dualism within shame—need not be. Exactly what, if anything, unifies shame of types 1 and 2 remains obscure. But clarifying the distinction between the two types does not on its own offer any comfort to the would-be defender of the morality system who claims to have discovered a form of distinctively moral shame.

7.

The indeterminate boundaries of the morality system bring me, finally, back to guilt. Survivor guilt might be thought to be a problem for any theory of guilt. If guilt is—as I think it often is—a victim-focused emotion, who is the survivor's victim supposed to have been? If guilt is—as is also often said—all about transgression, or disobeying the internal figure of the "enforcer," superego, etc.,[20] the same problem arises in a different form: since it is not apparent what if anything the survivor has *done*, it's not clear what transgression the enforcer or superego is supposed to be latching on to. All this might be thought to play well for enemies of the morality system: friends of morality have tended to say that guilt is a more purely moral emotion than shame, and survivor guilt is thus an embarrassment for the relationship between guilt and morality, because of the obscurity of its relationship to intentional action and thence to responsibility.

The tenacity of the association between guilt and morality comes out in an explanation sometimes offered for survivor guilt: as a psychoanalyst colleague once said to me, "There must be a fantasy of harm."[21] To the extent that there is only a *fantasy* of harm, of course survivor guilt is irrational. But at least a fantasy *of harm* successfully recruits guilt to the familiar model: guilt goes with responsibility for intentional wrongs. However, the "fantasy of harm" explanation is unsatisfactory. To stand a chance of being true, the fantasy has

---

20 See perhaps unexpectedly Williams 1993, and psychoanalysis *passim*.

21 For what it's worth, the same explanation finds its way in the Wikipedia article on survivor guilt (Wikipedia 2023): "A mental condition that occurs when a person believes *they have done something wrong* by surviving a traumatic or tragic event when others could not" (my italics). For a different psychoanalytic perspective on survivor guilt not dissimilar to my own, see Modell et al. 1983.

to be unconscious, because in many cases of survivor guilt there is certainly no *consciousness* or indeed outward record of harm anywhere to be seen. But that makes it notoriously difficult to evidence: the main evidence for the fantasy seems to be the prior conviction that guilt always has to go with responsibility for intentional wrongs. If there's an alternative explanation of survivor guilt, that conviction should lapse.

Here, then, is an easy case of survivor guilt: I am sitting in the back seat of a car beside a friend when our driver, who unbeknownst to us is drunk, goes off the road. Just thanks to the angle at which we leave the road, the obstacles that happen to be in the way, etc., my friend dies and I survive. I feel guilty. Something similar is also becoming apparent as a consequence of the Covid-19 pandemic, where vulnerable people who survived hospitalisation feel guilty knowing that others, seemingly no unhealthier than themselves, did not. It's important to this example that the survivors and the non-survivors both received the same level of hospital care. There are of course other examples where one might think—and possibly truly—that one's receiving hospital care had deprived someone else of it, as could be the case even if hospital beds were assigned by lottery, and still more so if by sharp-elbowed competition. To be sure, in either case there would be no *intended* harm, but to relate guilt to causing harm unintentionally—as in Williams' (1981) case of the bus driver who runs over someone who strays into the path of the bus—would be to aim a blow against the morality system at a different point.[22]

I want to suggest that there is an alternative explanation of survivor guilt that has nothing to do with harm, intentional or otherwise. On this view, the roots of survivor guilt connect with the idea of desert, in particular with discomfort at any unequal distribution of outcomes which is not grounded in desert, where grounding them in desert requires them to be grounded not in (for example) undeserved contingencies such as physical strength, but in the "purely moral capacity" (Williams 1973, 234) of effort. Moreover, survivor guilt is not a *sui generis* kind of guilt, but rather a species of a wider genus, namely guilt at the undeserved. Another—arguably commoner—member of the same genus is guilt at being richer or having a bigger garden than other people (and so not about surviving anything); this is betrayed by the fact that this guilt is typically allayed by remarks such as "I worked hard for it," i.e., those which try to connect the outcome with effort and thus with desert after all. And we find an echo of the same priorities in the converse thought, that in an ideal world order *nothing* would be undeserved—as in the on and off modern discomfort with competitive sport or competitive examinations, which of course do reward effort but also reward merely natural and therefore unearned differences in talent or physique.

The proffered explanation abandons any attempt to connect survivor guilt with responsibility for intentional harms. All the same, it comes not from outside the morality system but from a different corner of it, and so to that extent the association between guilt and the morality system is reinforced—though we should also perhaps conclude, as I suggested earlier, that the morality system is not much of a system. Nonetheless, sceptics about the association should not worry: there is reason enough from other quarters to think that the distinctive association between guilt and morality has been exaggerated.[23]

---

22 This thought supplies a reason to steer well clear in this investigation of one very well documented case of what's called survivor guilt, namely the guilt experienced by survivors of Nazi concentration camps. The reason is that in such environments—as in any where survival depends in part on competition for scarce resources—it may be very hard to tell to what extent one's own survival was or was not at another's expense, or therefore whether even inadvertent harms to others were caused. This is not to offer any opinion on whether any given concentration camp survivor did or did not inadvertently harm others, but simply to say that, to focus the present puzzle posed by survivor guilt, it is better to look at cases where this difficult empirical question clearly doesn't arise.

23 See for example my 2007.

References

Aubrey, J. 1957. *Brief Lives*. Secker and Warburg.

Deigh, J. 1983. "Shame and Self-Esteem: A Critique." *Ethics* 93 (2): 224–45.

Harcourt, E. 2007. "Guilt, Shame, and the 'Psychology of Love'." In *The Academic Face of Psychoanalysis*, edited by L. Braddock and M. Lacewing, 133–47. Routledge.

———. 2013. "Happenings outside one's Moral Self: Reflections on Utilitarianism and Moral Emotion." *Philosophical Papers* 42 (2): 239–58.

———. 2016. "Moral Emotions, Autonomy and the 'Extended Mind'." *Phenomenology and Mind* 11, 100-12.

Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge University Press.

Maibom, H. 2010. "The Descent of Shame." *Philosophy and Phenomenological Research* 80 (3), 566-94.

Modell, A., J. Weiss, and H. Sampson. 1983. *Narcissism, Masochism and the Sense of Guilt in Relation to the Therapeutic Process: Bulletin of the Psychotherapy Research Group, Dept of Psychiatry, Mount Zion Hospital and Medical Center #6*. June 1983. http://cmtcenter.net/wp-content/uploads/2017/07/Modell_Weiss_Sampson1983.pdf

Putnam, H. 1973. "Meaning and Reference." *Journal of Philosophy* 70 (19): 699–711.

Russell, B. 1927. *Why I Am Not a Christian*. Watts.

Shakespeare, W. 1608/1997. *King Lear*. Ed. R.A. Foakes. Bloomsbury.

Sophocles. 2008. *Electra and Other Plays*. Translated by D. A. Raeburn and P. E. Easterling. Penguin.

Wikipedia. 2023. "Survivor Guilt." Wikimedia Foundation. Last modified 8 October 2023, 17:59. https://en.wikipedia.org/wiki/Survivor_guilt

Williams, B. 1973. "The Idea of Equality." In his *Problems of the Self*, 230-49. Cambridge University Press.

———. 1981. "Moral Luck." In his *Moral Luck*, 20-39. Cambridge University Press.

———. 1985. *Ethics and the Limits of Philosophy*. Fontana.

———. 1993. *Shame and Necessity*. University of California Press.

Wollheim, Richard. 1983. *The Mind and Its Depths*. Harvard University Press.

———. 2004. *Germs: A Memoir of Childhood*. Waywiser.

A Strange State of Mournful Contentment
The Role of Compassion in Moral Betterment

Laura Candiotto - University of Pardubice

Abstract

In this paper, I will consider a unique case in which changing one's character is part of a process of moral betterment when facing oppression. By engaging with the Dutch-Jewish intellectual and Holocaust victim Etty Hillesum, I will highlight the situated dimension of moral betterment as a practice that is driven by the pressure of concurrent events. I will claim that moral betterment does not need to come out of an internal will to change for the better. Instead, I will argue that "bearing real suffering" (Hillesum 1996: 220) is what makes compassion a potential source of moral betterment. This is possible because, in compassion, one experiences emotional friction between weakness and strength in facing the suffering caused by oppression.

Keywords: moral betterment; compassion; Etty Hillesum; emotional friction; suffering; oppression.

1. Introduction

Why should I change my character for the better? What can support me in this process? And is merely changing my character what I am aiming for?

In this paper, I will consider a specific case in which changing one's character is part of a process of moral betterment in facing oppression. I take moral betterment as an "itinerary" into compassion to (1) transform moral egoism into altruism, and (2) avoid the trap of hating one's oppressors. By engaging in a dialogue with the Dutch-Jewish intellectual and Holocaust victim Etty Hillesum (1996), I will highlight the situated dimension of moral betterment as a practice that is driven by the pressure of concurrent events. I claim that moral betterment does not just come out of an internal will to change for the better; I argue instead that "bearing real suffering" (Hillesum 1996: 220) makes compassion a potential source of moral growth. This is possible because, in compassion, one experiences emotional friction. The emotional friction inherent to compassion can take many different shapes, for example as a tension between one's vulnerability and concern for others, the desire to relive suffering and the doubt about one's ability to do so, or the wish to help one's

friends and the guilt of disfavouring others. In its basic terms, I take the emotional friction inherent to compassion to be a tension between weakness and strength while facing suffering. As Etty[1] states, this is not just "an idea of suffering" (Hillesum 1996: 220) but a personal, intimate meeting with a fellow human being who is suffering. Importantly, this suffering is both personal and political. It comes out of an unbearable situation, the misery of being confined in Westerbork, the constant fear of being sent to Auschwitz, and, in Etty's case, the compassion towards what was happening to the others.[2]

To unpack how moral betterment can emerge out of emotional friction, I present two processes of transformation *of* and *through* compassion from the experience of Etty Hillesum. The aim here is neither descriptive nor simply historical. I take Etty Hillesum as a moral exemplar, in line with a core motif of ancient Greek ethics and contemporary virtue exemplarism (Zagzebski 2017). Her writing expresses such a rich and dense examination of the complexity of inner life and its transformations that I cannot reduce it to a mere "case" to analyse. Instead, I dare to philosophically encounter Etty's personal experience through her autobiographical reasoning,[3] learn from her, and bring part of what has emerged from this meeting to readers. This is because a moral exemplar does something to those who are inspired by it, be it through moral exhortation, reliance in moral development, or a desire to emulate. But, I will pursue this aim with the awareness that reading and writing are not enough to fully engage with Etty's life experience. As Etty insightfully expressed, "I shall never be able to put down in writing what life itself has spelled out for me in living letters. I have read it all, with my own eyes, and felt it with many senses" (Hillesum 1996: 210).

In the second section, I will introduce the two core concepts that will be employed in this paper, namely compassion and moral betterment. I will also introduce the metaphor of an "itinerary" and explain why it is best suited to helping us grasp the ethical significance of Etty's experience in Westerbork. In section 3, I will explain why we need to understand moral betterment as a situated phenomenon, in this case both in Etty's personal experience and in the historical situation, by diving into the affective dynamics of compassion. I will then explore Etty's diaries and letters to depict two instances of situated moral betterment in and through compassion. In *Moral Betterment 1*, there is a change intrinsic to compassion, namely from self-gratification and a parochial interest in relieving the suffering of a limited number of family members to an extended love of humanity. In *Moral Betterment 2*, compassion is seen as embedded in a set of practices that enabled Etty not to hate her oppressors. In both cases, love emerges as a power that orients compassion to everyone in a very embodied manner, as an intimate meeting with the "concrete other."

2. Compassion and Moral Betterment

Though philosophers have been interested in compassion for centuries, whether compassion is an emotion or a virtue is still under debate, and there is ongoing disagreement about its specific differences to related

---

1 In this paper, I mostly refer to Etty Hillesum as "Etty." However, I need to stress that this is not a way to diminish her value as a philosopher (who are usually referred to by their surnames), nor is it meant as a diminutive way of addressing her because she is a young woman. On the contrary, calling her by her first name is part of the methodology I will employ in this paper, which is to philosophically engage with her personal experience in an intimate way and to treat her as a moral exemplar. It also makes sense to address her in this way because she signed her diary and letters with a simple "Etty."

2 Etty volunteered to work as a social worker in the transit camp of Westerbork, the Netherlands, where Jewish people were held before being sent to the extermination camp in Auschwitz, Poland.

3 On the importance of narrative in developing what has been called a "science of personal wisdom," see Ferrari, Weststrate, and Petro 2013.

notions such as empathy and love, and its role in morality (Carr 1999; Nilsson 2011; Maibom 2014; Bloom 2016; Betzler 2019). I am not going to tackle this issue here. Instead, I will focus on how the affective dimension of compassion, which, as I claim, is inherently ambivalent and tensional, can contribute to moral betterment. I do not argue that this affective dimension of compassion is the sole source of moral betterment. There are many other factors, some that are not necessarily affective, which can ground moral betterment, such as religious beliefs or personal values. It is very often the case that they are all in play, and the quality and composition of the mix depends on the moral agent's personal history, as well as the culture and society in which she is living. But I will claim that when the affective dimension of compassion is involved, moral betterment assumes a situated shape: it takes the shape of an ethical itinerary that guides one through suffering with other beings and the desire to relieve them from it. I will first provide a conceptual clarification of the key terms I will be using, and then spell out my thesis using the "itinerary" metaphor.

In suffering together with another being, which is the experience implied by the etymology of the word "compassion,"[4] the agent feels both the other's suffering and the wish to relieve it. This action-oriented dimension of compassion as wanting to do something to relieve the other's suffering is also affectively experienced. Compassion does not only involve suffering: there is a desire to relieve suffering, accompanied by a sense of resilience and self-efficacy when someone believes that she is capable of accomplishing this task. One can feel good when helping others and find satisfaction through this act when the wish to relieve suffering is fulfilled, even if a successful result cannot be guaranteed. The felt dimension of compassion is not just an afflicted suffering; it implies a feeling of agency, a fundamental "I can." But there is also a sense of vulnerability that arises out of the acknowledgment of suffering and the wish to relieve it. One might also be afraid of hurting others while trying to help them. Or, one might feel incapable of helping others, for different reasons, personal and contextual. For instance, one might feel the urgency to help but not know what or how to do it. Or, one might know what to do but be in a situation in which one cannot. In this case, one's feeling of agency is tarnished and mixed with vulnerability. This means that compassion's affective dimension is far from simple. In compassion, there is a tensional ambivalence[5] between strength and weakness. I argue that this tension, instead of putting agency on hold, contributes to one's inherent motivation to strive for betterment. I will explain how moral betterment can unfold from this tensional ambivalence through the term "emotional friction" in the following section.

By moral betterment, I mean a type of moral change that is oriented toward the good. In moral betterment, one becomes more virtuous, similar to how one becomes more evil in moral deterioration. I am not referring to the moral motivation to change, as most of the literature on moral enhancement does (Douglas 2008; Jotterand and Levin 2019).[6] Instead, I am targeting character transformation through the application of certain practices, which would result in having virtue as a well-entrenched character trait. Pierre Hadot has identified and listed many spiritual exercises that were part of philosophical life in antiquity, such as the Stoic view from

---

4   The English word "compassion" comes from the Latin "compassion," from the verb compati (*com-pati*), which in turn derives from the Greek "*sym-pathein*," both meaning "to suffer together with." The German "Mitleid" has the same etymology (*mit-leiden*), which is where the Dutch "medelidjen" (the word employed by Etty in her diaries and letters) comes from.

5   Valence is the technical term for the feeling dimension of an emotion. It is usually understood in polar terms: pleasure/pain (Russell 1980, Frijda 1986) or good/bad (Kahneman 1999). Notably, Colombetti (2005) has challenged the standard, dichotomous conceptualisation of valence, showing that emotional experience is much more complex than the classical contrasting of polarised hedonic feelings implies. I take the affective dimension of compassion as a prime example of this more complex hedonic feeling by understanding the complexity in a dialectical manner, as will become clearer when I discuss "emotional friction" in the next section.

6   Here, I am taking a classical virtue ethics approach to moral betterment because it is the one that best expresses the moral framework of Etty Hillesum.

above and *praemeditatio malorum*, or the Epicurean *tetrapharmakos* (Hadot 1995: 81–125).[7] He claimed that these ancient spiritual exercises aimed at the transformation not only of one's character, but also of one's lifestyle. The reason for this is that spiritual exercises possessed "a concrete attitude, a way of life and seeing the world" (Hadot 1995: 108).[8] This reference to moral betterment in ancient Greek and Roman ethics is crucial, because it shows how moral betterment can come out of continuous practice. In the case we are considering, we will focus on compassion, and how it entails both a transformation of character and of lifestyle.

By moral change, I mean the transformation of moral character over time. This type of character transformation implies the renegotiation of moral frameworks, mindsets, character traits, and habits (Hämäläinen 2017). In moral betterment, one can attend to an improvement of moral character. As argued by Nora Hämäläinen (2017), different metaphors can be used for understanding moral change beyond the culturally laden metaphor of "progress." For example, moral change can be understood as a "tipping point," namely a change of balance within a framework of understanding. Another metaphor is that of a "bargaining table," which is particularly useful because it treats moral change as something that one does, as a practical negotiation between different moral frameworks.

The metaphor that I will use in this paper to shed light on the role of compassion in moral betterment is that of an "itinerary." Moral betterment, in this sense, is a path to virtue. This is an ancient metaphor of a noble philosophical breed.[9] The path is, to quote Rickard Kraut, "a route of goodness" (Kraut 2007: 15). According to Kraut, following this route is a legitimate method for attaining virtue for most ancient philosophers. Ancient ethics has an intrinsic teleological aim, with the good as its final end, and recognises that becoming virtuous is what enables the moral agent to achieve the good. This final end, namely attaining happiness as the good life (*eudaimonia*), motivates one to engage in spiritual exercises. Notably, for Aristotle, the good life consists of human flourishing: completely actualising one's potential, fully being what one is (Baracchi 2008: 79–101). Through this process of development and flourishing of potentialities, the subjective experience of happiness opens up a good life, which transcends the individual self (Hadot 1995).

The metaphor of an "itinerary" shares with this perfectionist tradition a fundamental developmentalist and teleological view on *eudaimonia* and virtue. However, it brings out something more specific about the role of the subject who will take the journey of moral betterment. The itinerary metaphor brings out the active role of the person who is walking the path, and who will commit herself to certain practices. It orients the development of the character around steps and practices that one does to move forward on the path. This is because the

---

7   Hadot (1995: 126–44) has also stressed the continuity between Ancient Greek philosophy and Christian philosophy (and notably, Ignatious of Loyola's *Exercitia Spiritualia*) regarding spiritual exercises. In this context, it is also important to note how much the metaphor of an "itinerary" is present in the spiritual and mystical traditions, and how it also took the shape of the practice of pilgrimage to holy places since the Middle Ages. See Bowie 1995 for a reading of Etty Hillesum's inner life through the lens of female mysticism.

8   But there is no contrast between the two. The transformation of the style of life comes from the transformation of how the subject perceives reality. For Hadot, this proto-cognitive account of transformation derives not only from Plotinus, heir of Plato's account on the conversion of the soul (Hadot 1993), but also from French phenomenology and existentialism (Hadot 2001). For me, it is important to stress that the transformation of the style of life is not simply a transformation of perception. What changes is the daily practices of a person, what she does, as well as how and why she does what she does. That is why, as Michel Foucault insightfully claimed, it is not just a matter of "care of the soul" but also of "ergon." See Candiotto 2021.

9   Parmenides understood the method of inquiry as an *odos*. He pointed to two paths, the paths of Night and Day (DK B1, in Graham 2010: 211). The former is one of not-being and cannot be followed because not-being is not; the latter is of the being, and that should be followed because it is the path of truth (DK B6, in Graham 2010: 215). Plato built his dialectical method as an ascending path. Just think about the way out of the cave in the *Republic* (*Resp.* 7, 514a-517a, in Cooper 1997: 1132–34) or the ladder of love in the *Symposium* (*Symp.* 210a-e, in Cooper 1997: 492–3).

itinerary metaphor stresses the personal commitment of walking that path by undertaking certain steps. The transformations do not simply happen of their own accord. Some effort, planning, and moral strength are required to reach the final destination.[10] Lastly, this itinerary is "ethical"[11] because it guides the transformation of character for the better, and is the path by which this transformation unfolds.

3. Situated Moral Betterment

In Etty's experience, this ethical itinerary is situated in her life, which is constantly burdened by the violence of her time. Situating her ethical itinerary within her personal life, which is so existentially bound to Jewish history, is important to avoid treating moral betterment in an abstract and unworldly manner. Here, "situated" means that Etty's moral betterment unfolds out of, and is rooted in, (1) her personal experience, and (2) the historical situation. Arguably, these two scales, the personal and the historical, are always related. But in Etty's moral betterment, they are so deeply entangled that her personal life can exemplify what a "counter scenario" of extermination can be (Gaarlandt 1996: xiii), in particular with regards to the reciprocal beliefs of victims and perpetrators.[12] Etty chose to join the Jewish fate by volunteering at Westerbork, and in so doing, she opened her heart to the suffering of humanity (1) and, at the same time, to the risk of falling into the trap of hating her oppressors (2).

In this section, I will first contextualise Etty's personal experience within the historical situation of the occupied Netherlands. Then, I will explain how moral betterment can come out of an engagement with real suffering as the result of "emotional friction."

In May 1940, the German army invaded the Netherlands, which, until then, was a neutral country with respect to the war. In the spring of 1942, the Nazis began their first major roundups of Jews in the Netherlands. Until the Nazis' pressure on Jews intensified in 1942, Etty's personal life did not change much.[13] She was studying at the university, taking part in some intellectual circles in Amsterdam, and enjoying herself with friends and lovers, most remarkably the psychochirologist Julius Spier. But then she realised that she could not continue the life she was living.[14] After becoming a member of the Jewish Council on the 15th of July, 1942, she volunteered as a social worker at Westerbork, a Nazi transit camp in the occupied Netherlands. From the end of July 1942, Etty lived and worked in Westerbork. Throughout the fall of 1942, she intermittently spent time in Amsterdam

---

10 Scholnick (1994) has analysed the "path metaphor" in comparison with the "arrow metaphor" from an embodied perspective (Johnson 1987), by stressing that, in the path metaphor, there is not just linearity and directedness but also the experience of passing through different steps. She argued that understanding a path with the help of the arrow metaphor makes the path too deterministic. This does not allow any change from the walker. I agree with the criticism of teleology as an arrow, especially from an embodied perspective. However, in the path metaphor there is more than this. The path metaphor understands change as progress and growing. So, I take the "itinerary metaphor" as a subspecies of the path metaphor, with the key characteristic of planning the steps for making the growing happen. This is relevant to Etty's strength of character and act of will in undertaking a process of moral betterment as an itinerary in and through compassion. So, there is a teleological tension in an itinerary – it is not just wandering. But this tension is not the linearity of the arrow, but rather the wilful intention to better oneself through the journey.

11 The adjective "ethical" derives from the Greek word *ethos*, which means character.

12 Heidi Maibom (2022: 86-105) has recently analysed the asymmetry of perspectives between victims and perpetrators. The former tends to moralise transgressions and to see the wrong as incomprehensible, while the latter tends to justify their actions. Etty offered an extraordinary alternative to this binarism by committing herself to compassion and avoiding the temptation of hatred.

13 "Last night I wondered again if I was so unworldly simply because the German measures affect me so little personally" (1996: 129; Diary, 29 April 1942).

14 "Only a few months ago I still believed that politics did not touch me and wondered if that was 'unworldliness,' a lack of real understanding" (1996: 126; Diary, 29 April 1942).

due to her poor health. But this was also an opportunity for her to collect medicine to bring to Westerbork and to collaborate with the leftist student resistance front. From June 1943, Etty lived permanently in Westerbork, although some friends offered her the opportunity to escape. Her parents and her brother Misha were also sent to the transit camp during the summer. In September 1943, the entire family was deported to Auschwitz. Etty wrote her last postcard from the train: "We left the camp singing" (Hillesum 1996: 360). In November 1943, Etty was sent to the gas chamber.

In her diaries and letters, Etty expresses a deep, nuanced, and complex emotional life. With constant introspection that was mediated by the practice of writing in a diary,[15] she vividly depicts the transformations in her inner life that were prompted by the external changes to her lifestyle that took place upon her move to Westerbork. This existential change significantly marks her moral betterment because it allowed her to meet "reality" and "bear real suffering":

> Reality is something one shoulders together with all the suffering that goes with it, and with all the difficulties. And as one shoulders them, so one's resilience grows stronger. But the idea of suffering (which is not the reality, for real suffering is always fruitful and can turn life into a precious thing) must be destroyed. And if you destroy the ideas behind which life lies imprisoned as behind bars, then you liberate your true life, its real mainsprings, and then you will also have the strength to bear real suffering, your own and the world's. (1996: 220; *Diary*, 30 September 1942).

In this diary entry, we find Etty's self-exhortation to destroy the mere idea of suffering because the strength to bear real suffering can only be found in liberating life from the bars of ideas.[16] Etty was sceptical about relying on ideas because she noticed in her first-person experience that, in doing so, there was a high risk of experiencing compassion in a narcissistic manner, namely as "self-gratification" (1996: 10; *Diary*, 9 March 1941). This practical imperative to meet reality is what facilitates a transformation of compassion, as we will see in the next section. Here it is important to highlight that the situatedness of Etty's moral betterment is chosen by Etty as a practice to liberate her true life. So, stressing the two scales of situatedness is not to make a general claim about the personal and historical location of an experience. On the contrary, my emphasis on the entanglement between Etty's personal experience and the historical situation in which it was situated points to the moral significance of situatedness. It is by making her commitment to freeing herself from ideas a practice of meeting real suffering that Etty's moral growth can unfold. This is because, for her, a mere idea of compassion is an abstraction – she needs to immerse herself in the situation to make it real. She does not want to live in fantasies or in writing about compassion from her safe desk in Amsterdam. She felt the urgency to dig deep into real suffering in order to avoid the narcissism implied by the idea of compassion. But what does it take to meet real suffering? How is the itinerary of compassion a source of moral betterment?

My answer lies in what José Medina (2019) calls "emotional friction." Medina distinguishes between negative and positive emotional friction (Medina 2019: 26). The former refers to the emotional obstacles one must face to overcome one's complicity in oppression; the latter refers to the emotional attitudes that help one to resist oppression. Medina takes emotional friction as something that should be cultivated and sustained in social and political activism. Here, I intend to apply this conceptualisation to moral development. In doing so, I interpret emotional friction in a very embodied and embedded manner, since I assume that the "friction"

---

15 This can also be understood as a practice of resistance. On this, see Brenner 1997.
16 This criticism towards a mere idea of suffering that obscures the access to reality resonates with Murdoch (1992: ch. 18) and Weil (2003, in particular the section titled "Imagination which Fills the Void").

Medina is talking about results from experiencing oppression in a very intimate manner, namely by actively engaging with the suffering of a person who is a victim of oppression and all the ambivalence that comes with it. This engagement with the other's suffering creates friction not only because oppression is against one's personal values, but also because the suffering is conjoined with the desire to relieve it. As I have already said in the previous section, there is more than suffering to compassion. There is also the desire to relieve suffering, and reliance on one's strength to do so.

Let us dig a bit deeper into the dynamics of emotional friction in compassion. Compare the vulnerability felt when acknowledging another person's suffering, and the toughness felt when desiring and committing to do something to relieve it. The relationship seems clear-cut: vulnerability is unpleasant, and toughness is pleasurable. Vulnerability discloses weakness, and toughness discloses strength. Vulnerability and toughness are polar opposites and seem irreconcilable. But, in fact, this is not always the case: the rich and nuanced affective dimension of compassion is a good counterexample.

Some relevant affective dispositions come with compassion. For example, accepting that one will undergo suffering means to accept feeling vulnerable,[17] and this is not refuted by the desire to free others from suffering by employing one's resilience.[18] The desire to relieve suffering activates and keeps alive the dialectic between weakness and strength, but it does not remove their opposing natures. This is because there is dynamism and tension in this ambivalence, and it is precisely this tensional ambivalence that characterises my understanding of emotional friction within an embodied and embedded framework. Emotional friction disrupts and challenges the character, like a spark lit by rubbing two wooden sticks together. Emotional friction is morally significant in that it can disrupt character and serve as a source of moral betterment in an embodied and embedded manner – as we will see by engaging with Etty's experience.

Now consider the world-directedness of affective valence. In the case of compassion, an affective engagement with suffering produces a tension between vulnerability and resilience. This tension is not simply "internal." The real suffering is acknowledged in the others' suffering and by engaging with it daily. Compassion could be expressed, in Etty's words, as "a strange state of mournful contentment":

> Yes, children, that's how it is, I am in a strange state of mournful contentment. If I once wrote you a desperate letter, don't take it too much to heart; it expressed only a brief moment. It's true you can suffer, but that need not make you desperate. And now I'm going to jump in at the deep end again; I'm off to the hospital with a little tin box for my beloved father under one arm and my official folder under the other. I shall find many empty beds there after this transport. (1996: 288; *Letters*, dated 29 June 1943)

There is mourning in compassion, especially if it is experienced in a transit camp where, every week, hundreds of inmates are sent to Auschwitz. But there is also a sort of contentment that comes from being at the service of others. Etty had a little box for her father under one arm and an official folder under the other. This scene significantly expresses Etty's daily search for balance in the ambivalence between the care for her parents and for all the inmates. As we will see in the next section, this tension will mark the transformation of parochial compassion into something bigger. But here compassion is lived as a "strange feeling," namely complex, ambivalent, and tensional. It is ambivalent because one can feel good (contentment) while feeling bad (sorrow for someone's suffering and eventual death). There is a temporality at work here: it is not that this affective

---

17 On vulnerability, especially in facing illness, see Brady 2018: 103–11.
18 On resilience as a virtue of the character that comes out of suffering, see Brady 2018: 88–102.

ambivalence should necessarily make one feel good and bad at the same time. Affective ambivalence is not necessarily a contradictory feeling, if by "contradiction" we mean, in Aristotelian terms, being one thing and its negation at the same time and in the same respect. The ambivalence might be extended through time, in a dialectical process of ups and downs, alternations of balance and tension, desperation and peace.

> Yes, really, it's true, there are compassionate laws in nature, if only we can keep a feeling for their rhythm. I notice that afresh each time in myself: when I am at the limits of despair, unable, I am sure, to go on, suddenly the balance shifts over to the other side, and I can laugh and take life as it comes. After feeling really low for ages, you can suddenly rise so high above earthly misery that you feel lighter and more liberated than ever before in your life. I am now very well again, but for a few days I was quite desperate. Equilibrium is restored time and again. Ah, children, we live in a strange world. (1996: 305; *Letters*, dated 8 July 1943)

It is not that just our emotional life is "strange"; the world is. And Etty experienced this fact through immersing herself in the fate of her people. But her choice to engage with real suffering should not be taken as martyrdom. Meeting the suffering of her time was still an "unelected transformative experience" (Carel and Kidd 2020: 167). She was simply ridding herself of the privileged unworldliness she described in the first pages of her diary. In Etty's experience, fighting against the abstraction of compassion and immersing herself in real suffering was the key to making compassion a source of moral betterment.[19] It is also important not to apply a reductionist frame onto the nature of this strange feeling. In my understanding, it certainly expresses the affective dynamics of compassion, but there is something more there:

> This morning, while I stood at the tub with a colleague, I said with great emotion something like this: "The realms of the soul and the spirit are so spacious and unending that this little bit of physical discomfort and suffering really doesn't matter all that much. I do not feel I have been robbed of my freedom; essentially no one can do me any harm at all." (1996: 227–8; Letters, dated 29 June 1943)

These lines come immediately before the preceding quotation. They talk about a profound sense of freedom that can be experienced, even in Westerbork. The itinerary metaphor can help illuminate how this is possible. As in an itinerary, there are many steps and pauses that make up the route, so the transformation of and through compassion does not happen by its own accord and all of a sudden. There are also other feelings that arise along the way. And feelings are not the only things that matter: freedom, acceptance, love, and also compassion, are all more than feelings. Existential meanings, commitments, and values are also part of an ethical itinerary.

The dialectical dimension of compassion's ambivalence also emerges when we consider the freedom that Etty was referring to in the above quotation. There is not just acceptance in freedom; there is also resistance. Etty's desire to relieve suffering marks both her existential resistance and her acceptance. It is a form of resistance because her ethical itinerary contrasts with the itinerary of the trains which arrived every Monday evening in Westerbork and left the following day with hundreds of people destined for Auschwitz. Etty was not delusional. She was acutely aware that she was in an extermination camp. She also knew that her survival

---

19 It might be argued that only distal compassion is a virtue, and not proximal compassion, since it requires the exercise of rational capacities. For this argument, and debate about it, see Bagnoli 2018. Although I cannot argue for it here, I would like to stress that Etty's compassion is not merely "proximal," although being situated in the real personal encounters, since it is framed within an itinerary of personal development orientated by love.

meant someone else's death. Therefore, it is clear she accepted the fact that "what they are after is our total destruction" (1996: 154; *Diary*, 3 July 1942). But she was also strong in reclaiming an inner freedom that cannot be stolen. She didn't fall into desperation, although she was honest enough to acknowledge her vulnerability. For example, she frankly admitted: "I am still too afraid of hurting others, when it actually hurts them much more if I am with them in fact but not at all in spirit" (1996: 116; *Diary*, 22 April 1942). But in Westerbork, she worked to "find life meaningful" (1996: 154; *Diary*, 3 July 1942) by working with her soul, "or whatever else you care to call what shines through from within" (1996: 29; *Diary*, 8 June 1941). It is important to stress that this is not an intimist refuge when one gives up on political hope.

> If we were to save only our bodies and nothing more from the camps all over the world, that would not be enough. What matters is not whether we preserve our lives at any cost, but how we preserve them. I sometimes think that every new situation, good or bad, can enrich us with new insights. But if we abandon the hard facts that we are forced to face, if we give them no shelter in our heads and hearts, do not allow them to settle and change into impulses through which we can grow and from which we can draw meaning –then we are not a viable generation. It is not easy – and no doubt less easy for us Jews than for anyone else –yet if we have nothing to offer a desolate postwar world but our bodies saved at any cost, if we fail to draw new meaning from the deep wells of our distress and despair, then it will not be enough. New thoughts will have to radiate outward from the camps themselves, new insights, spreading lucidity, will have to cross the barbed wire enclosing us and join with the insights that people outside will have to earn just as bloodily, in circumstances that are slowly becoming almost as difficult. And perhaps, on the common basis of an honest search for some way to understand these dark events, wrecked lives may yet take a tentative step forward. (1996; 250; Letters, dated 18 December 1942)

This quotation comes from one of the two letters that were illegally published by the Dutch Resistance in 1943. This is important to stress because it shows how much Etty's ethical itinerary was not simply personal therapy. It was an act of acceptance of and resistance to oppression; it was a message for the future.

4. Moral Betterment 1

In the previous sections, I argued that emotional friction is experienced in the tensional ambivalence of compassion. According to Medina (2019), emotional friction is what activates change. Change takes place both in promoting positive emotional friction – in Etty's case to work in and through compassion to extend it to humanity (*Moral Betterment 1*) – and combating negative emotional friction – in Etty's case to counter the tendency to hate oppressors (*Moral Betterment 2*). In this section, I will examine some pages from the *Diaries* and the *Letters* to shed light on these two specific types of moral betterment. I am considering the *Diaries* and the *Letters from Westerbork* together in order to focus on the ethical itinerary undertaken by Etty, where she strives to apply her aspiration for compassion to the reality of suffering. In this process, love will emerge as the power that orients compassion towards everyone. But let me once again repeat that this reconstruction of the dynamics of moral betterment should not be narrowly conceived in an abstract, mechanical, or reductionist manner. As has been stressed by Eva Hoffman:

> The process of change is always mysterious, but in the middle stages of Etty's diary, it becomes almost palpable. . . . Increasingly, her writing points toward that border where attentiveness to subjectivity meets contemplation, where emotional intuition converges with moral thought. In all of this, Etty had

the courage to follow the thread of her own experience – and that thread took her further still, in more unexpected and radical directions. (Hoffman 1996: xii)

The two processes of moral betterment I am describing here are *personal*. They are *Etty*'s moral betterment; they are *her* ethical itinerary. But this does not mean that we cannot relate to or learn from them. We can develop our own sensitivity, be inspired by her experience, and learn what we should not do if we take the practice of reading her writing as one important step within our own ethical itineraries. But we should avoid making a universal theory out of it because, in doing so, we would betray her warning against imprisoning life behind the bars of ideas (Hillesum 1996: 220).

> Wednesday night. . . . My protracted headaches: so much masochism; my abundant compassion: so much self-gratification. Compassion can be creative, but it can also be greedy. . . . For example, clinging to one's parents: one has to see them as people with a destiny of their own. (1996: 10–11; *Diary*, 9th March 1941)

> When I suffer for the vulnerable, is it not for my own vulnerability that I really suffer? (1996: 230–1; *Diary*, 13th October 1942)

Here, Etty is warning herself against an overly narrow form of compassion that aims at self-gratification, namely to be the person who can help the miserable by virtue of being in a safer and privileged position, or to care only for one's closest neighbours, in Etty's case her parents, even if in a selfless manner.[20] This is a form of compassion that is actually a form of self-interest, because it is centred around one's own suffering. Etty was looking at an alternative to moral egoism, or the duty to act in one's own interest. But she was very careful to avoid construing compassion as altruistic *per se*. She understood that even compassion can be egoistic:

> What does wise mean to you: egoistic? All this egoism is getting so boring. Since people have been telling each other for centuries that man is basically an egoist, one begins to believe it and actually becomes egoistic. There are so many sides to a human being that it would be nice to try something else, just for a change from boring and unproductive egoism. (1996: 262; *Letters*, dated 8 April 1943)

> We used to feel so certain that we would help each other bear the sorrows of our age. . . . Vanity of vanities – but what was not vain was my discovery that I was able to commit myself unreservedly to another, to bind myself to him, and to share his sorrow. (1996: 222; *Diaries*, 30 September 1942)

By immersing herself in the dynamics of compassion, she found the possibility to commit herself *unreservedly* to another. This absence of restrictions is a crucial shift in her moral betterment as it marks the passage from moral egoism to altruism through love.

> Last night I had to struggle again not to be overwhelmed by pity for my parents, since it would paralyze me if I gave in to it. I know that we must not lose ourselves so completely in grief and concern for our families that we have little thought or love left for our neighbors. More and more I tend toward the idea that love for everyone who may cross your path, love for everyone made in God's image, must rise above love for blood relatives. (1996: 334; *Letters*, dated 18 August 1943)

---

20 This coincides with her biggest fear, which is not directed at herself but rather to her loved ones: "The hardest labor camp is better than this suspense every week. It didn't bother me so much before, because I had accepted the fact that I would be going to Poland. But living in fear for your loved ones, knowing that an infinitely long martyrdom is in store for them while your own life here stays relatively idyllic, is something few can bear" (1996: 297; *Letters*, dated 5 July 1943).

Here, Etty pushes herself to extend compassion to everyone. This "everyone" is not an abstract other, but everyone "who may cross your path." In Etty's itinerary, unreserved compassion is always about particular beings with whom she is sharing her life and fate. In Etty's experience, parochial compassion is not transformed into a universal disposition that risks becoming blind to the particular needs and qualities of a particular person she meets. On the contrary, it is always an intimate encounter with a concrete other. This is made possible by the power of love:

> Why not turn the love that cannot be bestowed on another, or on the other sex, into a force that benefits the whole community and that might still be love? And if we attempt that transformation, are we not standing on the solid ground of the real world, of reality? A reality as tangible as a bed with a man or a woman in it. (1996: 208; *Diaries*, 20 September 1942)

Here Etty finally finds the reality she sought by immersing herself in real suffering. She feels that love can be a force directed towards the whole community, and yet be as intimate and personal as the way it is experienced in a romantic encounter.[21] I say that she "feels" this because the reality she is referring to is "tangible," i.e., it is not imprisoned behind the bars of ideas. It is not abstract, but concrete, namely situated in her daily life experience. In my understanding, this is a powerful alternative to two degenerations of love: love as self-interest and love as universal.[22] Although opposite, these two types of love share an important feature, namely that they both leave out engagement with a concrete other. Etty's aim, on the other hand, was to meet reality. In the tangible encounter with another human being, Etty discovers love as an unrestricted force that is due to everybody. By helping that specific concrete other, her compassion can take part in this broader force of love and find the strength to help another human being.[23] Therefore, love is also situated in Etty's itinerary. It shapes her itinerary in a manner that does indeed extend the threads of compassion but, at the same time, it takes place in the particular encounters with those who cross her path; therefore, her path is not simply one of suffering. In meeting suffering, she found love.[24] And with love, beauty, meaning, and acceptance come:

> When one has once reached the point of experiencing life as something significant and beautiful, even in these times, or rather precisely in these times, then it is as if everything that happens has to happen just as it does and in no other way. (1996: 200; Diaries, 15 September 1942)

Before moving on to the next section, I want to highlight another important feature of Etty's itinerary into and through compassion. Emotional friction is not only the trigger of transformation, and does not operate just

---

21 It might be argued, as for example by Hoffmann (1996: x), that in doing so, Etty made a transition from *eros* to *agape*. If it is true that her experience of love at the camp was characterized by the other-oriented attitude that is typical of *agape*, at the same time she never denied the sensual dimension that is typical of *eros*. For Etty this embodied dimension of love is important because it makes the meeting with the other real, concrete, and tangible. The kind of love she is looking for is like the one that is experienced in a bed with a woman and a man in it. So, if we were to take *agape* as a selfless attitude that contrasts with *eros*, as Hoffmann (1996: x) does, this would dismiss the deeply embodied dimension of her love and compassion.

22 For a discussion of these two degenerations, see Candiotto and De Jaegher 2021.

23 This is a powerful alternative to the much-debated ethical limits of empathy as directed only to those nearby. In Etty's experience, helping a concrete other made it possible to extend compassion to everyone, and did not result in parochialism.

24 Although Etty's first commitment was to the extension of love to humanity, it is important to notice how sensitive she was to the beauty of nature, even in something as small as the grounds in Westerbork, and how aware she was of the healing power of nature. For instance, "My red and yellow roses are now fully open. While I sat there working in that hell, they quietly went on blossoming. Many say, 'How can you still think of flowers!' Last night, walking that long way home through the rain with the blister on my foot, I still made a short detour to seek out a flower stall, and went home with a large bunch of roses. They are just as real as all the misery I witness each day" (1996: 188; *Diary*, 23 July 1942).

at the beginning of the itinerary. Being able to extend compassion does not mean that Etty no longer feels the tension between her vulnerability and her desire to care for everyone. Later in her journey, in 1943 when her parents were sent to Westerbork, she found herself overwhelmed by pity for them. Every day, she also faced the tension between her (limited) power to postpone her parents' departure for Auschwitz and the awareness that, by doing so, she would have sent someone else who deserved the same compassion that she devoted to her parents (and herself – because she wanted to have them alive).[25] For me, this is not a failure of her ethical itinerary. On the contrary, it demonstrates how moral betterment is a constant practice that needs to be kept alive on a daily basis. The tensional ambivalence in emotional friction is precisely what feeds the itinerary.[26] An ethical itinerary is made of ups and downs; it is not linear progress. There may be roundabouts, points where one has to turn back, or has to stay longer. There may be uphill and downhill sections. Moments when one goes faster and others when one trudges. However, these diversions actually contribute to the process of transformation through emotional friction. But we can also observe a discipline in it – Etty does not let herself be overwhelmed by emotional friction. Etty is often as severe with herself as she is kind to others. She is aware that she needs to undertake this itinerary seriously if she wants to resist the temptation of hating the oppressors, as we will see in a moment. Even if the real itinerary is made while walking a path, the destination is fixed. There is a *moving towards* in her ethical itinerary, although very often the practices she engages in have to do with a *moving away from*, in particular from moral egoism. In both movements, her discipline is nourished by the power of aspiration.

Aspiration is an important force in Etty's ethical itinerary. It is what gives her the strength to continually find meaning in a degraded life in Westerbork. Early on, and before definitively moving to the camp, Etty wrote that she wanted to be "the thinking heart of the barracks" (Hillesum 1996: 225). This aspiration – or prayer (as she depicts it on the same page) – helped her maintain her commitment in voluntarily joining the Jewish fate, even though she was given an opportunity to escape. This aspiration supported her when a narrower form of compassion grew in her heart, especially in meeting the sorrow of her parents. And this aspiration of acting as "a balm to all the wounds" (1996: 231; *Diary*, 11 October 1942) was always there, even on the last page of her diary.

5. Moral Betterment 2

Etty's ethical itinerary, explicitly understood as guided by compassion, helped her keep her hatred for Nazi officials at bay, even in the face of the enormous suffering they inflicted on their victims, including herself. For Etty, to keep the distinction between oppressor and victim does not necessarily mean to love the victim and hate the oppressor. This would have tarnished her soul and her passionate recognition of the meaning of life. How could she resist the shortcut of indifference and the temptation to hate the oppressors?

---

25 Etty was not involved in selecting who to send to Auschwitz, but she could put pressure on those who had the power of making the list since she was working for the Jewish Council. But, with the worsening of the situation, this privilege was lost. She tried to remove her parents from the list even on that Monday in September when the whole family was sent to Auschwitz, but without success.

26 I cannot delve into the much-debated problem of complicity in the oppression of the Jewish Council's members here. On this, see Arendt 1963 and a discussion in Lederman 2018. Especially in the Netherlands, the "well organised" Jewish Council contributed towards having the highest number of Jewish victims in Western Europe. But I would like to suggest that the emotional friction Etty experienced in compassion might be also be what helped her avoid complicity and maintain an indignant stance, all the while not hating her oppressors.

In this section, I depict three crucial steps that mark her ethical itinerary: 1. the thinking heart; 2. allowing oneself to be a battlefield; and 3. believing in humanity.

Etty engages in the constant work of introspection by writing her diary. The ethical itinerary comprises a constant relation with oneself, others, the world, and all of the problems and goodness that come from them. Writing a diary and sending letters to her acquaintances helped her do the inner work of reflecting on the experience. Etty is not just a woman of feelings. She wants to be the *thinking* heart of the barracks. This means that, by exploring her heart through writing a diary, she aims to better understand her choices, commitments, and aspirations. In fact, her commitment to constant introspection is what allows her ethical transformation to take place.

> At night, as I lay in the camp on my plank bed, surrounded by women and girls gently snoring, dreaming aloud, quietly sobbing and tossing and turning, women and girls who often told me during the day, "We don't want to think, we don't want to feel, otherwise we are sure to go out of our minds," I was sometimes filled with an infinite tenderness, and lay awake for hours letting all the many, too many impressions of a much-too-long day wash over me, and I prayed, "Let me be the thinking heart of these barracks." And that is what I want to be again. The thinking heart of a whole concentration camp. I lie here so patiently and now so calmly again, that I feel quite a bit better already. I feel my strength returning to me; I have stopped making plans and worrying about risks. Happen what may, it is bound to be for the good. (1996: 225; Diaries, end of September 1942)[27]

By aspiring to be the thinking heart of a whole concentration camp, Etty committed herself to avoiding two emotional reactions: indifference and hatred.

> For those who have been granted the nerve-shattering privilege of being allowed to stay in Westerbork "until further notice," there is the great moral danger of becoming blunted and hardened. The human suffering that we have seen during the last six months, and still see daily, is more than anyone can be expected to comprehend in half a year. No wonder we hear on all sides every day, in every pitch of voice, "We don't want to think, we don't want to feel, we want to forget as soon as possible." It seems to me that this is a very great danger. (1996: 250–1; *Letters*, dated 18th December 1942)

Indifference is here understood as a great moral danger because it obstructs compassion. Etty is aware that indifference is a common response to all the suffering that is around. Still, she thinks it is a danger she needs to avoid. And she does it by allowing herself to be a battlefield. She does not close the door of her heart to avoid suffering. On the contrary, she allows emotional friction to take place:

> I feel like a small battlefield, in which the problems, or some of the problems, of our time are being fought out. All one can hope to do is to keep oneself humbly available, to allow oneself to be a battlefield. . . . In that respect, I am probably very hospitabl; mine is often an exceedingly bloody battlefield, and dreadful fatigue and splitting headaches are the toll I have to pay. (1996: 32; *Diaries*, 15 June 1941)

Avoiding indifference is one thing, but how can one avoid leaving this battlefield full of hatred?

---

27 Notably, in mid-September she would already sign the diary as "the thinking heart" (Hillesum 1996: 199).

Yesterday afternoon we read over the notes he had given me. And when we came to the words, "If there were only one human being worthy of the name of 'man,' then we should be justified in believing in men and in humanity," I threw my arms round him on a sudden impulse. It is the problem of our age: hatred of Germans poisons everyone's mind. "Let the bastards drown, the lot of them" – such sentiments have become part and parcel of our daily speech and sometimes make one feel that life these days has grown impossible. Until suddenly, a few weeks ago, I had a liberating thought that surfaced in me like a hesitant, tender young blade of grass thrusting its way through a wilderness of weeds: if there were only one decent German, then he should be cherished despite that whole barbaric gang, and because of that one decent German it is wrong to pour hatred over an entire people. That doesn't mean you have to be halfhearted; on the contrary, you must make a stand, wax indignant at times, try to get to the bottom of things. But indiscriminate hatred is the worst thing there is. It is a sickness of the soul. Hatred does not lie in my nature. If things were to come to such a pass that I began to hate people, then I would know that my soul was sick and I should have to look for a cure as quickly as possible. (1996: 11; Diaries, 15 March, 9:30)

Etty chose to believe in humanity despite everything. But this was not a result of a universalist imperative. As in the case of love just presented in the previous section, this commitment came from engaging with a concrete other. Etty says to herself that, if there were even one single decent German, she would not have to hate all of them. This resolution arose as a liberating thought. This thought is liberating because it goes beyond the dualism of loving the victim and hating the oppressor. This thought takes the shape of an ethical commitment that appreciates *difference*. What is wrong is *undifferentiated* hatred. It might be objected that it is wrong to hate all the Germans but that, as one might love the one decent German that one meets, at the same time one should hate the one evil German one meets. And, unfortunately, there were many, not just one, at the camp. Although Etty discovered at the end of her journey that it was extremely hard to find even this one decent German at the camp – she noticed that the camp guards transformed too, unfortunately for the worst[28] – she strongly claimed that hatred was always bad because it would have tarnished her soul. Why is this the case? Because it would make one like the oppressor:

Take one of my colleagues. I see him often in my thoughts. The most striking thing about him is his inflexible, rigid neck. He hates our persecutors with an undying hatred, presumably with good reason. But he himself is a bully. . . . Klaas, all I really wanted to say is this: we have so much work to do on ourselves that we shouldn't even be thinking of hating our so-called enemies. We are hurtful enough to one another as it is. (1996: 210–1; *Diaries*)

At the same time, she stresses that avoiding hate does not mean refusing to take a stand or express moral indignation.[29] It is not the case that as we love one concrete other we should hate another concrete other. There

---

28 "On earlier transports, some of the guards were simple, kindly types with puzzled expressions, who walked about the camp smoking their pipes and speaking in some incomprehensible dialect. One would have found their company not too objectionable on the journey. Now I am transfixed with terror. Oafish, jeering faces, in which one seeks in vain for even the slightest trace of human warmth. At what fronts did they learn their business? In what punishment camps were they trained?" (1996: 349; Letters, dated 24 August 1943).

29 Already before going to Westerbork, Etty had a clear opinion about the difference between hatred and indignation: "But genuine moral indignation must run deep and not be petty personal hatred, for personal hatred usually means little more than using passing incidents as excuses for keeping alive personal hurts, perhaps suffered years ago. Call it psychology, but we can't let ourselves be led astray any longer; we must look at all that indignation we feel and discover whether its roots are genuine and deep and truly moral" (1996: 130; *Diary*, 29 April 1942).

is no equivalence between love and hatred.[30] Hatred, for Etty, is always a sickness of the soul (Hillesum 1996: 11), and so it should be cured as quickly as possible.[31] But this does not mean that one should be impartial:[32]

> The absence of hatred in no way implies the absence of moral indignation. I know that those who hate have good reason to do so. But why should we always have to choose the cheapest and easiest way? It has been brought home forcibly to me here how every atom of hatred added to the world makes it an even more inhospitable place. And I also believe, childishly perhaps but stubbornly, that the earth will become more habitable again only through the love that the Jew Paul described to the citizens of Corinth in the thirteenth chapter of his first letter. (1996: 255–6; *Letters*)[33]

Etty stood against the oppression in the camp with the power of an extended form of compassion. This form of compassion is beyond the opposing logic of loving the victims and hating the oppressors because it is unreservedly extended to all humanity. This means that sometimes Etty was even capable of feeling compassion for the guards. But this did not mean turning a blind eye to the horror and misery or disavowing the violence perpetuated by the oppressors. Still, compassion moved by love was what could help build a "whole new world" after the war:

> All I wanted to say is this: The misery here is quite terrible; and yet, late at night when the day has slunk away into the depths behind me, I often walk with a spring in my step along the barbed wire. And then time and again, it soars straight from my heart – I can't help it, that's just the way it is, like some elementary force – the feeling that life is glorious and magnificent, and that one day we shall be building a whole new world. Against every new outrage and every fresh horror, we shall put up one more piece of love and goodness, drawing strength from within ourselves. We may suffer, but we must not succumb. And if we should survive unhurt in body and soul, but above all in soul, without bitterness and without hatred, then we shall have a right to a say after the war. (1996: 294; *Letters*)

6. Conclusion

In this paper, I stressed that there is more to compassion than suffering, through engaging with Etty's writings. There is resilience and vulnerability, aspiration and presence, weakness and strength. Far from being made of just a simple, polarised feeling, the affective dimension of compassion is characterised by a tensional ambivalence in emotional friction. I took the emotional friction inherent to compassion to be a tension

---

30 This contrasts one important equivalence in the virtue-theoretical discourse about good and evil, namely that if x is intrinsically evil, hating x for itself is intrinsically good. See Hurka 2001: 16.

31 There might be religious reasons in support of this choice. For example, it might be argued that Etty needed to cure her soul and avoid hatred for an afterlife salvation. There is a debate about the religious character of Etty's faith, if Jewish, Christian, or neither. I cannot go deeper into it here (see Coetsier 2014: 71–123 for references). But I would like to stress that my reading supports an understanding of Etty's commitment to take care of her soul as a practice relevant to her life in the camp and under the pressure of the history. In this manner, I tend to read her spiritual journey alongside the ancient model of care of the soul, as I have already explained.

32 Impartiality seems to be a core characteristic of ethical judgment. However, as has been skillfully remarked by Maibom (2022: 220–47), impartiality in ethics presupposes an ideal observer who can rightly judge a situation by being outside of it. But moral agents are actors that hold specific perspectives, not indifferent spectators. And this is what can make them care more about others, not less, as in Etty's moral experience that is so much rooted in her personal encounter with others' suffering.

33 See Adinolfi 2011: 49–50, 96–97, 135, 168 for the important role played by Saint Paul's exhortation to reply to hatred with love in Etty's ethical itinerary.

between weakness and strength while facing suffering that is caused by oppression. I argued that emotional friction is experienced in meeting a concrete other who is suffering from oppression. Then, I showed how this emotional friction kindles moral betterment by analysing two types of moral betterment in and through compassion. Emotional friction can be simply destructive. But in Etty's experience it was constructive, because she found love through bearing real suffering. In Etty's ethical itinerary, meeting with suffering didn't take the shape of martyrdom. On the contrary, her ethical itinerary unfolds as a passionate aspiration to a greater and more profound form of compassion, one that is oriented by the love for every concrete other she meets in her journey.

References

Adinolfi, I. 2011. *Etty Hillesum: La Fortezza Inespugnabile*. Il Melangolo.

Arendt, H. 1963. *Eichmann in Jerusalem: A Report on the Banality of Evil*. Viking Press.

Bagnoli, C. 2018. "Compassion and Practical Reason: The Prospective of the Vulnerable." In *The Moral Psychology of Compassion*, edited by C. Price and J. Caouette, 125–45. Rowman and Littlefield.

Baracchi, C. 2008. *Aristotle's Ethics as First Philosophy*. Cambridge University Press.

Betzler, M. 2019. "The Relational Value of Empathy." *International Journal of Philosophical Studies* 27 (2): 136–61.

Bloom, P. 2016. *Against Empathy: The Case for Rational Compassion*. Ecco.

Bowie, F. 1995. "Modern Women Mystics: Etty Hillesum and Simone Weil." *New Blackfriars* 76 (892): 175–87.

Brady, M. 2018. *Suffering and Virtue*. Oxford University Press.

Brenner, R. F. 1997. *Writing as Resistance: Four Women Confronting the Holocaust: Edith Stein, Simone Weil, Anne Frank and Etty Hillesum*. Pennsylvania State University Press.

Candiotto, L. 2021 "Care of the Self and Politics. Michel Foucault: Heir of a Forgotten Plato?" In *Platonism: Ficino to Foucault*, edited by V. Rees, A. Corrias, F. M. Crasta, L. Follesa, and G. Giglioni, 297–313. Brill.

Candiotto, L. and H. De Jaegher. 2021. "Love In-Between." *The Journal of Ethics* 25: 501–24.

Carel, H. and I. Kidd. 2020. "Suffering as Transformative Experience." In *Philosophy of Suffering*, edited by D. Bain, M. Brady, and J. Corns, 165–179. Routledge.

Carr, B. 1999. "Pity and Compassion as Social Virtues." *Philosophy*, 74: 411–29.

Coetsier, M. G. S. 2014. *The Existential Philosophy of Etty Hillesum: An Analysis of her Diaries and Letters*. Brill.

Colombetti, G. 2005. "Appraising Valence." *Journal of Consciousness Studies* 12 (8–10): 103–26.

Cooper, J. M., ed. 1997. *Plato: Complete Works*. Hackett Publishing Company.

Douglas, T. 2008. "Moral Enhancement." *Journal of Applied Philosophy* 25 (3): 228–45.

Ferrari, M., N. M. Weststrate, and A. Petro. 2013. "Stories of Wisdom to Live By: Developing Wisdom in a Narrative Mode."

In *The Scientific Study of Personal Wisdom*, edited by M. Ferrari and N. M. Weststrate, 137–64. Springer.

Fridja, N. 1986. *The Emotions*. Cambridge University Press.

Gaarlandt, J. 1996. "Introduction." In *Etty Hillesum, An Interrupted Life: The Diaries, 1941–1943, and Letters from Westerbork*, edited by J. G. Gaarlandt, xiii-xxi. Henry Holt and Company.

Graham, D. W., ed. and tr. 2010. *The Texts of Early Greek Philosophy: The Complete Fragments and Selected Testimonies of the Major Presocratics*. 2 vols. Cambridge University Press.

Hadot, P. 1993, *Plotinus, or The Simplicity of Vision*. Translated by M. Chase. Chicago University Press.

———. 1995, *Philosophy as a Way of Life: Spiritual Exercises from Socrates to Foucault*. Edited by A. I. Davidson, translated by M. Chase. Basil Blackwell.

———. 2001, *La Philosophie comme Manière de Vivre. Entretiens avec Jeannie Carlier et Arnold I. Davidson*. Albin Michel.

Hämäläinen, N. 2017. "Three Metaphors toward a Conception of Moral Change." *Nordic Wittgenstein Review* 6 (2): 47–69.

Hillesum, E. 1996. *Etty Hillesum, An Interrupted Life: The Diaries 1941–1943 and Letters from Westerbork*. Edited by J. G. Gaarlandt, translated by J. Cape, foreword by E. Koffmann. Henry Holt and Company.

Hoffman, E. 1996. "Foreword." In *Etty Hillesum, An Interrupted Life: The Diaries, 1941–1943, and Letters from Westerbork*, edited by J. G. Gaarlandt, vii-xii. Henry Holt and Company.

Hurka, T. 2001. *Virtue, Vice, and Value*. Oxford University Press.

Johnson, M. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason*. University of Chicago Press.

Jotterand, F. and S. B. Levin. 2019. "Moral Deficits, Moral Motivation and the Feasibility of Moral Bioenhancement." *Topoi* 38: 63–71.

Kahneman, D. 1999. "Objective Happiness." In *Well-Being: The Foundations of Hedonic Psychology*, edited by D. *Kahneman*, E. Diener, and N. Schwarz, 3–25. Russell Sage Foundation.

Kraut, R. 2007, *What is Good and Why: The Ethics of Well-Being.* Harvard University Press.

Lederman, S. 2018. "Hannah Arendt's Critique of the *Judenräte* in Context: Modern Jewish Leadership and Radical Democracy." *Holocaust and Genocide Studies* 32 (2): 207–23.

Maibom, H. L., ed. 2014. *Empathy and Morality*. Oxford University Press.

Maibom, H. L. 2022. *The Space Between*. Oxford University Press.

Medina, J. 2019. "Racial Violence, Emotional Friction, and Epistemic Activism." *Angelaki*, 24:4, 22-37.

Murdoch, I. 1992. *Metaphysics as a Guide to Morals*. Chatto & Windus.

Nilsson, P. 2011. "On the Suffering of Compassion." *Philosophia* 39: 135–44.

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

Scholnick, E. K. 1994. "Route Metaphors." *Psychological Inquiry* 5 (3): 254–56.

Weil, S. 2003. *Gravity and Grace*. Routledge.

Zagzebski, L. 2017. *Exemplarist Moral Theory*. Oxford University Press.

Epistemic Oppression and Affective Exclusion:
A Pragmatist Approach

Aleksandra Hernandez - University of Miami

Abstract

Epistemic oppression is the systematic exclusion of oppressed groups from knowledge-making practices. One of the goals of feminist epistemology is to understand the nature of the barriers preventing oppressed individuals from realising their epistemic agency and contributing to knowledge production. Feminist epistemologists have taken monumental strides in theorising the different forms of exclusions that constitute epistemic oppression (e.g., Fricker 2007; Dotson 2014; Bailey 2014). But more needs to be done by way of understanding what makes an oppressive epistemic landscape resilient and resistant to change. In this paper, I develop a pragmatist account of the affective mechanisms in place that protect the epistemic landscape from critical reflection and produce what I refer to as *affective exclusion*. I argue that, since the mechanism that protects the epistemic landscape from critical reflection is affective by nature, we cannot reason ourselves out of affective exclusion. Instead, I propose that, although they often prevent us from realising our epistemic agency, our co-occurring emotions and feelings, such as grief, anxiety, and other epistemic feelings, can be catalysts for reflection about the skewed nature of the epistemic landscape. I conclude with a discussion of the role that narrative fiction might play in helping us resist affective exclusion and ending epistemic oppression.

Keywords: Co-occurring emotions; affective exclusion; epistemic oppression; self-knowledge; grief

1. Introduction

Epistemic oppression is the systematic exclusion of oppressed groups from knowledge-making practices. One of the goals of feminist epistemology is to understand the nature of the barriers preventing oppressed individuals from realising their epistemic agency and contributing to knowledge production. Feminist epistemologists have taken monumental strides in theorising the different forms of exclusions that constitute epistemic oppression (e.g., Fricker 2007; Dotson 2014; Bailey 2014). But more needs to be done to understand what makes an oppressive epistemic landscape resilient and resistant to change in the first place. In this paper, I propose that what makes oppressive epistemic landscapes resilient and resistant to change are the affective

habits of individuals from dominant groups, whose emotional responses function to protect the epistemic landscape from critical reflection by way of *affective exclusion*. In this paper, I develop an account of epistemic oppression that specifies what affective exclusion is, how it is generated, how it is experienced, and how it can be resisted.

My argument will proceed as follows. I begin by introducing the concept of affective exclusion and the way it interacts with other forms of exclusions to produce epistemic oppression. Then, I offer a pragmatist account of co-occurring emotions to shed light on the affective processes involved in generating affective exclusions. Here, we will see that the affective practices of privileged individuals function as the safety valves of the epistemic landscape, insofar as they protect it from critical reflection. We will also see that we cannot reason ourselves out of affective exclusion, and that precisely that which prevents us from recognising ourselves as epistemic agents can enable us to become critics of the epistemic landscape: our emotions and feelings. An analysis of a short story by Joy Williams called "Shepherd" (2015) follows, which dramatises the role that grief and other emotions and feelings can play in initiating resistance to affective exclusion. There I argue that, although emotions can and frequently do distort our understanding of the world (Goldie 2004), they can also bring into focus what matters to us the most, and, in doing so, serve as catalysts for new ways of understanding ourselves and the world around us. I conclude with some preliminary thoughts on the significant challenges we face in stimulating critical reflection about the skewed nature of the epistemic landscape more generally, so as to enable oppressed individuals to become knowers and critics of the epistemic terrain.

2. Affective Exclusion

We know epistemic oppression exists and that to combat it we need to include individuals from marginilized communities in our knowledge-making practices; we need to take them seriously as epistemic agents and give them the conceptual tools to realise their epistemic agency. I shall refer to these types of exclusions as *standing exclusion* and *conceptual exclusion*.[1] Standing exclusion involves calling into question the credibility, status, or *standing* of oppressed epistemic agents as knowledge producers. Examples of standing exclusion include testimonial injustice, which occurs when the credibility of an oppressed individual is undermined, and when epistemic authority is automatically conferred on members of privileged groups (Fricker 2007). Standing exclusion relies on implicit bias and stereotyping. Conceptual exclusion, by contrast, involves making unavailable the conceptual resources necessary for oppressed epistemic agents to participate in knowledge-making practices. An example of conceptual exclusion is hermeneutical injustice, which occurs when individuals lack the hermeneutic skills, concepts, and language to *interpret* and *understand* their experiences and produce knowledge about themselves and the world (Fricker 2007).

Yet there is another type of exclusion more fundamental than either standing or conceptual exclusion—what Dotson (2014) refers to as "third-order" exclusion. Dotson first introduced this concept to address an undertheorised problem within the epistemic injustice literature concerning the resistance an epistemic agent might face from the epistemic landscape itself. This resistance can be described as follows: even when oppressed agents have access to the shared epistemic resources, the epistemic resources present a barrier to the agent, insofar as, in order for the agent to make her position legible—that is, in order for her position to

---

1 These types of exclusions roughly correspond to Dotson's first- and second-order exclusions. I do not stick with Dotson's terminology here because her typology of epistemic exclusions implies that these exclusions are hierarchical, and that they can occur in isolation from one another.

receive uptake—either the existing epistemic resources must be revised or new epistemic resources must be added to the epistemic landscape. According to Dotson, without adequate epistemic resources, "[the agent's] testimony . . . may be rejected as nonsensical; they [dominant knowers] may designate her as a deceiver with dangerous ideas; or the conclusions she draws might even invoke ridicule or laughter" (130). Dotson further claims that third-order exclusions are not reducible to social or economic forces, which suggests that the only way to resist them is by marshalling imaginative and affective resources and taking up *a bird's eye point of view*, as Alison Bailey claims.

There is a problem, however, with the way that the problem is formulated: to the extent that dominant epistemic landscapes reflect the habits of thought and practices of dominant knowers, it is not clear that third-order exclusions are irreducible to social or economic forces. Moreover, as standpoint feminists have argued, it is impossible to occupy an agent-independent perspective,[2] which suggests that, to specify the nature of this type of exclusion, and to resist it, we need to excavate the epistemic landscape itself from *within*, rather than from without, as only then will we have better tools at our disposal for combatting epistemic inertia and bringing about "new ways of knowing" (Bailey, 2014: 67). In other words, we need to draw lessons from our lived, affective experiences to understand what makes the epistemic landscape robust, resilient, and inert. This is because, even when the conceptual resources that enable oppressed individuals to acquire standing and become epistemic authorities are available, there is no guarantee that these resources will be used without the adequate *affective* tools. When an individual is unable to realise their epistemic agency due to self-undermining affective habits, for example, they are coming up against epistemic barriers that are affective in nature. Similarly, when an oppressed individual is not taken seriously by privileged epistemic agents because they haven't yet acquired the right language to articulate their knowledge, they can regress back to self-doubt. This is an affective barrier that can be unwittingly reinforced by privileged epistemic agents. I refer to this type of exclusion as *affective exclusion*.

Affective exclusion is a type of injustice that is reducible to social and economic forces, insofar as oppressed individuals acquire affective habits that undermine their confidence in their capacities as knowledge producers. It is also widespread and difficult to address, as it concerns the overall affective habits of the community of knowers. anxiety and guilt, I propose, generate self-doubt and are among the affective habits that function as safety valves that kick in when the epistemic landscape is under threat. Since privileged individuals are the beneficiaries of the unlevel playing field, their tendency to respond with anger when challenged or called out serves the function of protecting their standing, which in turn influences how conceptual resources are shared. Their affective habits also play an unwitting, critical role in preventing critical reflection about the skewed nature of the epistemic landscape, insofar as there's no reason—beyond the appeal of justice—for privileged epistemic agents to question their standing, and hence, to reflect on their privileged status. And, while it is in the interests of oppressed individuals to question the skewed nature of the epistemic landscape, without the right affective tools, oppressed individuals can remain locked in self-undermining affective loops which prevent them from seeking out the conceptual tools to articulate knowledge. We will see next that these affective loops are involved in a process I refer to as emotional foregrounding—a concept grounded in a pragmatist understanding of the inseparability of the mind and the body.

---

2 See Harding 1986 and Toole 2019. Dotson (2014) also discusses the importance of Standpoint Theory for conceptualising epistemic exclusions. Rogers (2021) critiques the notion that some forms of epistemic oppression are reducible to political power.

3. Emotional Foregrounding and Affective Exclusion

In this section, I develop an account of affective exclusion, anchored in the philosophy of William James and Charles Sanders Peirce, to (1) develop the concept of emotional foregrounding, a process in which co-occurring emotions come to the foreground from the flux of felt experience as discrete entities, and (2) argue that, while it is the selective mechanism that generates affective exclusions, emotional foregrounding can be co-opted to enable an agent to realise her epistemic agency. I do so by specifying the function of attending to and then abstracting emotions from the flux of felt experience: to help us evaluate the aptness of our emotional responses and change our affective habits. I also offer a pragmatist take on how we come to feel confident or certain about our attitudes toward a particular situation or problem, and end by enumerating the important ways that our emotions and feelings serve as important sources for envisaging new ways of knowing and inhabiting our worlds.

When William James famously wrote that "*bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur* IS *the emotion*" (James 1884: 189–90), critics were quick to dismiss him. Many objected to what they understood to be the main claim: that emotions reduce to bodily symptoms, and hence, that they are divorced from reason.[3] In an influential, alternative interpretation of James that seeks to set the record straight, Phoebe Ellsworth (1994) argues that, for James, an emotion is an awareness or a feeling of bodily changes, and that bodily changes follow one's perception of an *exciting fact*—the event, situation, or object perceived.[4] From this she concludes that (1) emotions are not the perception of one's bodily changes; (2) the perception of an exciting fact is an interpretation; (3) emotions are not discreet entities but rather processes that flow below the threshold of conscious experience; and (4) by the time we *catch* or perceive bodily changes, we are detecting that we are having an emotional experience that cannot be fully grasped or reproduced in its full complexity and totality.[5] On this reading of James, although emotions are co-occurring, we are not consciously *aware* of experiencing more than one emotion at a time, which suggests that our emotions are not always accessible to us through introspection. Even so, some emotions will eventually become the focus of our attention as we work out how best to feel about a situation. One might argue that emotions become the focus of our attention as the exciting facts come to the foreground. I propose, however, that we select emotions out of the flux of our felt experience using our epistemic feelings as guides.[6] The epistemic feelings I believe are involved in the process of emotional foregrounding are as follows.

---

3   For an in-depth look at the reception of William James's theory of emotions, see Ellsworth (1994), who gives a much fuller account of James's theory, which does not reduce it to the claim that emotions are perceptions of bodily changes, or worse yet, that they are the bodily changes themselves.

4   Jake Spinella (2020) has recently argued that Ellsworth omits instances where James appears to suggest that cognitive appraisals are not necessary for us to have an emotional experience. He also argues that James (1984; 1894) is a proto-functionalist theorist about emotions.

5   "We may catch," James writes, "the trick with the involuntary muscles, but fail with the skin, glands, heart, and other viscera" (James 1884: 192). Moreover, when we "fancy some strong emotion," we *abstract* "from our consciousness of it all the feelings of its characteristic bodily symptoms," leaving nothing behind but "a cold and neutral state of intellectual perception" (James 1884: 193). In other words, as soon as we catch ourselves becoming cognisant of or perceiving some bodily change, we are interrupting an emotional process and abstracting emotions as if they were discrete entities experienced in isolation from one another and from other mental states. For James, emotions are embodied, and thereby inextricable from our felt experience of bodily disturbances, and "without bodily sensations there could be no feeling of emotion" (Ellsworth 1994: 233).

6   See for example Nagel 2010 for an account of the central role epistemic anxiety plays in evidence gathering; and Vazard 2021 for the role epistemic anxiety plays in motivating us to reassess our epistemic states. See Kurth 2018a for a book length study of the intentional object of anxiety, which he conceives as an emotion.

Feelings of Doubt or Epistemic Anxiety: According to Christopher Hookway (2008), these feelings come up when we are contemplating beliefs we have already formed. They motivate inquiry by directing us to reassess our beliefs and to look for more evidence, and hence, they differ from Cartesian doubt, where we merely evaluate a belief as "unsafe" and become motivated to suspend it.

Feelings of Searching as a Species of Feelings of Knowing: According to Ronald de Sousa (2009), feelings of knowing are metacognitive feelings that indicate to us we know something before we can *retrieve* it from memory. In this paper, I conceive of the feeling that we are close to knowing something, or that we're on to something before we are able to articulate it, as a special kind of feeling of knowing, what I refer to as a *feeling of searching*. The concept of a feeling of searching was first articulated by Charles Sanders Peirce, who defined it as "a peculiar sensation belonging to the act of thinking that each of these predicates inheres in the subject. In hypothetic inference this complicated feeling so produced is replaced by a single feeling of greater intensity, that belonging to the act of thinking the hypothetic conclusion"(1878: 482).[7] Feelings of searching guide us through the process of inquiry by becoming "hot" or more intense when we are close to solving a problem. Feelings of searching in the sense I am developing here interact with epistemic anxiety and induce affective loops that are productive when they are rooted in feelings of self-trust. Affective loops are useful mechanisms for quickly solving problems, as we can cycle through various felt evaluations.

Feelings of Confidence or Certainty: According to de Sousa (2009), feelings of certainty end inquiry, as they indicate to us that we have solved the problem or have arrived at the answer we were looking for. Feelings of certainty are interlinked with the emotion of trust towards other persons whom we perceive as epistemic authorities. I would like to add that feelings of certainty about our beliefs require *self-trust* and regarding oneself as an epistemic authority.

To elucidate the role that epistemic feelings play in helping us navigate our epistemic landscapes, let me lead with an example. Suppose I am out all day and rushing to get back to my dog who desperately needs to go outside. When I get home, I see garbage all over the floor and experience a rush of feelings. I might react in anger because my dog broke the rules and rummaged through the trash. But I might also feel uncertain as to what to do next. I experience epistemic anxiety, in other words, which motivates inquiry not only into what happened and why it keeps happening, but also into the appropriate emotional response. I might find myself feeling that my anger is misplaced, and my guilt warranted, for I suspect that I have been leaving my dog alone for too long lately. It is because my attention oscillates between "Not again!" and "I shouldn't be leaving him home for so long," and "Poor boy, he is bored and perhaps even anxious" that it appears as if I were experiencing anger separately from guilt and other emotions.

What is in fact happening is that, in trying to make sense of my felt experience, I co-experience felt evaluations of the evidence. These felt evaluations form the basis of possible interpretations, which I navigate using my

---

7 According to Charles Sanders Peirce, "Hypothesis substitutes, for a complicated tangle of predicates attached to one subject, a single conception. Now, there is a peculiar sensation belonging to the act of thinking that each of these predicates inheres in the subject. In hypothetic inference this complicated feeling so produced is replaced by a single feeling of greater intensity, that belonging to the act of thinking the hypothetic conclusion. Now, when our nervous system is excited in a complicated way, there being a relation between the elements of the excitation, the result is a single harmonious disturbance which I call an emotion. Thus, the various sounds made by the instruments of an orchestra strike upon the ear, and the result is a peculiar musical emotion, quite distinct from the sounds themselves. This emotion is essentially the same thing as a hypothetic inference, and every hypothetic inference involves the formation of such an emotion" (1878: 482; 1931–5: 2.643; quoted in Trajkovski and Williamson 2021).

feelings of searching. One of the functions of feelings of searching is to foreground the emotions that best fit the working hypothesis. Thus, epistemic anxiety engages us in a process of affective looping, in which we cycle from one felt evaluation to another until we reach a feeling of confidence or certainty. In other words, epistemic anxiety, which is the engine of the affective looping process, interacts with feelings of searching until eventually we make up our minds and come to feel confident that we have understood the situation. When the process of inquiry ends and we feel certain that we understand the situation, our epistemic anxiety calms and other co-occurring emotions fade to the background. Epistemic anxiety is thereby an epistemic virtue, as it motivates the agent to look for more evidence, quickly formulate many working hypotheses at a time, and assess them using their feelings of searching as guides.[8] Good knowers trust themselves as epistemic authorities and become proficient at navigating epistemic anxiety.

Bad knowers, by contrast, avoid epistemic anxiety altogether and arrive at feelings of certainty very quickly. For example, although an entitled individual might feel certain about the aptness of their anger, their anger may be reinforcing their biases and perceptions about their standing. This is because their epistemic landscape is skewed in their favour, and initiates emotional responses that protect them from the epistemic anxiety and self-doubt that would inevitably arise, were they to engage in critical reflection about their privileged standing. Skewed epistemic landscapes thrive on the affective habits of bad knowers. An individual from an oppressed group experiences a higher level of uncertainty in similar situations, however, as they must grapple with conflicting ways of responding to a complex situation. In certain situations, an oppressed individual possesses epistemic virtues lacking in their more privileged counterparts, such as the tendency to doubt the rightness of their interpretations and the aptness of their emotional responses. The problem emerges when oppressed individuals neither trust themselves nor regard themselves as epistemic authorities. A lack of self-trust locks these individuals in self-undermining anxiety.[9]

For example, suppose that Hanna (1) has the hermeneutical resources, which in principle could enable her to produce knowledge, but (2) she does not regard herself as an epistemic authority. Hanna has a hard time producing knowledge because, although she is quite good at generating possible interpretations, her anxiety prevents her from trusting her searching feelings. Her anxiety is thereby not productive, as it is rooted in self-doubt. By contrast, suppose that another oppressed individual, Annie, (1) regards herself as an epistemic authority and (2) has the hermeneutical resources which enable her to produce knowledge. Annie frequently experiences epistemic anxiety, but is at the same time more resilient than her friend, Hanna, who does not view herself as an epistemic authority. Annie's feelings of searching interact productively with epistemic anxiety until she reaches the feeling of confidence or certainty about how to best understand her situation. Annie's emotional responses are bound to be more apt, where Hanna's will be skewed, due to the biases and false narratives that she has internalised, much like her privileged counterparts, which affect her perception of her own standing. Annie is thereby better equipped to cope with affective exclusion, but not because Hanna does not have the capacity to form better interpretations of her situations and engage in inquiry. Rather, what makes Annie resilient and a reliable epistemic authority is that her affective looping mechanism is rooted in self-trust. Hanna, on the other hand, remains locked in disabling anxiety, and comes to feel certain—if she ever does!—only when she accepts the false narratives endorsed by her privileged counterparts.

---

8   See Cabrera 2021 for an account of epistemic anxiety as an intellectual virtue.
9   See Kurth 2018b, Levy 2016, Maibom 2005, and Roskies 2003 for analyses of cases of anxiety in which individuals' ruminations and questioning of their beliefs are unreasonable and unproductive.

Peter Goldie argues, in "Emotion, Feeling, and Knowledge of the World," that sometimes "one's emotional feelings tend to *skew the epistemic landscape* to make it cohere with the emotional experience" (2004: 171), such that "when new evidence does emerge, one tends not only to be insensitive to the evidence but also for the sake of internal coherence, to doubt the reliability of the source of that evidence" (172). Significantly, Goldie claims that being in the middle of an emotional episode prevents us from taking a "dispassionate" perspective and evaluating the evidence more objectively (174), such as when I feel despair with my new job and fail to take seriously my friend's evaluation that work is not all that bad. Even more worryingly, he argues that sometimes we have no introspective access to our emotions, and so have no way of judging their appropriateness or how they might be distorting our interpretations of the evidence (175). For example, when I am unaware that my understandable but inappropriate anger at my colleague for missing an important meeting (I later find out her child was very ill) influences the letter of reference they have asked me to write after the incident, I fail to realise that my emotions are distorting my judgements about their personal qualities.

Now assuming that, in the first example, my friend's evaluation that my work is not as bad as it seems is not distorted by some other external factor, such as implicit bias, and that in the second example, it is my lingering anger and not some other additional factor such as sexism that is influencing how I approach the letter, I agree with Goldie that emotions can distort our beliefs and how we evaluate a situation. But these examples are set up in highly idealised conditions which do not adequately account for the reality that external factors, such as sexism and racism, do in fact influence our felt evaluations of the world, often unconsciously. I am not sure, in other words, that the mechanism that skews our epistemic landscapes is purely internally derived. If I am an entitled man, for example, I will look for reasons to justify my inappropriate burst of anger at work in response to a request by my woman colleague to involve more women in committee work. My reasons will also likely cohere with my pre-existing sexist beliefs and attitudes, which are lurking in the background, such as when I conclude that my female co-worker was wrong to have challenged me in front of my other colleagues, either because she did so inappropriately (in a confrontational, aggressive manner that does not correspond to my belief that women should be docile, fragile, and gentle) or because she must be wrong that not enough women are assigned committee work, since *she* has a committee assignment (thereby failing to take her seriously as an epistemic agent).

If, by contrast, I am a woman who has an angry outburst at work, I might be more likely to have difficulty feeling confident that my anger is apt. Suppose that I felt ignored and not taken seriously by my boss, and that, even after getting angry and pointing out to him my reasons for believing that I have been mistreated, he accuses me of being unreasonable and aggressive. I thus come to feel guilty and ashamed, and justifiably so—or so I believe—since I should have controlled my anger. Significantly, the reasons with which I justify my guilt and shame ultimately cohere with the false and oppressive narrative that I have internalised, and which caused me to feel guilty and ashamed in the first place: that women are not supposed to be aggressive. Here we see that an external factor (the stereotypes and implicit biases that I have internalised) distorts the way that I evaluate the evidence, such that I come to believe that I was wrong to respond to my boss's actions with anger, and hence come to feel guilt and shame. In contrast to the entitled male, whose anger coheres with his background sexist beliefs, the angry woman's internalised sexism does not cohere with her anger, which is why her emotional response shifts to guilt and shame. In both instances, it is the internalised oppressive norms that skew the respective interpretations of the evidence; however, because the entitled man experiences his anger as apt, his anger coheres with and reinformces the other attitudes and beliefs he holds. But the anger is not the distorting factor in this case; rather, it is an index of an already skewed epistemic landscape.

What I am suggesting is that, in these more complex kinds of social situation, it is not emotions that are misleading, but the implicit oppressive social norms that have conditioned us to feel the way we do. If I have been conditioned to believe that I should have more privileges than women because I am a man, I am bound to react with anger when a woman challenges the norms and practices that give men unearned privileges.[10] These beliefs are influenced by factors external to the agent, such as the entitled man's internalised structural sexism. His anger is thereby working as it should, insofar as it reliably tracks a transgression within the context of the structural sexism encoded in the epistemic landscape. What makes the emotion a reliable index of the nature of the epistemic landscape are the norms and practices that shape the sexist man's beliefs and attitudes.

Things get more complicated, however, when we are forced to grapple with an interpretation of events that does not cohere with the implicit biases we have internalised. Thus, suppose that (1) my friend witnessed my boss's sexist behaviour, that she insists that he is not taking me seriously, and that I take her judgement seriously and come to understand that I was a victim of sexism, but also that (2) a part of me still believes my boss in saying that I may have over-reacted. This causes me anxiety and distress, and has the effect of backgrounding my apt anger and of foregrounding my unjustified feelings of guilt and shame. Unlike the entitled man, whose anger indicates a skewed epistemic landscape, and unlike the angry woman, whose internalised false narrative distorts her evaluations of her behaviour, and causes her to perceive her anger as inapt, I co-experience emotions that correspond to two conflicting interpretations of the evidence: a true one, in which I am a victim of sexism and thereby justified in feeling angry, and a false one, in which I feel guilty and ashamed for not controlling my apt anger.

Although at first glance it might seem that dispassionate deliberation should settle the matter—that, all things considered, my friend must be right, and that eventually I will come to the right conclusion—we cannot ignore one aspect of my felt experience which explains why I oscillate between feeling anger and feeling guilt and shame: my epistemic anxiety. Whereas my entitled counterpart is bound to ignore the evidence because his sexist beliefs cohere with his anger, I am anxious to grapple with multiple interpretations of the evidence and solve the problem—it matters to me that I get this right! Yet some of these interpretations do not cohere with the feeling that he is right. This is because my anger was apt and because a part of me suspects that I was justified in expressing it; yet, my feeling that my anger was apt is at odds with my belief that I should have controlled my anger—a belief shaped by the stereotypes and implicit biases that I have internalised. Consequently, I experience epistemic anxiety as I try to work out whether to trust the feeling that I should have controlled my anger, or my feeling that my anger was justified—and no amount of dispassionate deliberation will guarantee that my emotions will align with the facts.

This brings me to Williams' story, which in many ways parallels the story of the anxious woman who struggles with the feeling that she over-reacted, and her feeling that her anger was apt. But there is one additional layer that complicates my analysis: the woman in this story is grieving the death of her dog. Throughout the story, the protagonist has flashbacks of the past, intermingled with her impressions of the present, as she tries to understand what happened. Grief, in this case, is a catalyst for inquiry into the events leading up to the death of her non-human companion. But grief also presents her with an occasion to assess her relationship with her fiancé; it gives her clarity of mind by crystallising what really matters to her, and it will remain the aim of the next section to lay out the mechanism by which grief can initiate critical reflection.

---

10 See Bailey 1998 for an analysis of male privilege and Manne 2017 for an analysis of how sexism reinforces traditional gender norms

4. Grief as a Catalyst for Inquiry and Critical Reflection

According to Michael Cholbi (2022: 79), grief is "a privileged epistemic route to our pasts" and an occasion for inquiry into the nature of the loss of a relationship that formed part of our practical identities. When we lose someone that matters to us, we are perplexed, confused, and disoriented, and in grappling with that inner turmoil, we come to see ourselves and our relationship with the person we have lost in a different light. The loss of someone we care about also initiates a change in focus, through which we are forced to determine how we will relate to them in the future, since we can no longer depend on that relationship for support, for accomplishing certain goals, for future planning, etc. For these reasons, grief is an important source of knowledge, insofar as it disrupts our self-narratives and our sense of self (81). In Cholbi's words: "Grief catalyzes what Quassim Cassam has called 'substantial self-knowledge,' knowledge of our 'values, emotions, abilities, and of what makes one happy.' As animals that live in community with other humans, our self-understandings and practical identities depend in large measure on the social world we share with others" (2022: 84–85).

Williams' "Shepherd" captures precisely what Cholbi has in mind when he argues that "the good in grief . . . is self knowledge" (2022: 83). Told from the perspective of a woman grieving the loss of her companion animal, the story presents fragments of her memories, intertwined with a whirl of co-occurring emotions and feelings, including happiness, love, regret, sadness, depression, anxiety, guilt, and shame. The dog, who was left alone out on the porch, breaks the screen, runs out, falls over a cliff, and then drowns. The woman's grief is thereby stalled by her feelings of guilt and shame for having left her German Shepherd alone many times over the course of her relationship with her fiancé, Chester—feelings that are amplified by the fact that he blames her for what happened and dismisses her grief. Over the course of their exchanges after the death of the Shepherd, it becomes very clear that her fiancé does not care about her dog, and hence, does not take on her cares in the way that a truly loving partner would. He thereby does not care for his fiancé as a person, where to care about someone as a person is to care about the things that matter to them (Helm 2010). The woman's German Shepherd is the exciting fact that stimulates her co-experienced emotions throughout the story. At the same time, she cares about her relationship with Chester, and it is one of the interesting aspects of the story that we are encouraged to piece together how the protagonist feels about him before and after the Shepherd's death.[11]

Unsurprisingly, the woman does not appear to have introspective access to her emotions—or at the very least, she does not reflect on them—leaving it up to readers to infer how she might be feeling, based on her flashbacks as well as her actions and reported thoughts. What we can infer for sure is that she was much happier with the Shepherd before she met her fiancé, and that she became acutely aware of this fact after she lost her dog. She says, for example, that she felt "light," and as if she had no other cares in the world when she first got him. She describes her relationship with her dog as mutually rewarding:

> When the girl and the shepherd had first begun their life together, they had lived around Mile 47 in the
> Florida Keys. The girl worked in a small marine laboratory there. Her life was purely her own and the

---

11 According to Helm (2009), emotions are structured by matters of import, or the things that we care about and on whose behalf we are prepared to act. Helm refers to the matter of import of an emotion as the *focus* of the emotion. For example, suppose that I care about my tomato plant and that there is a seventy percent chance of frost. When I hear the weather report predicting a seventy percent chance of frost tomorrow, the frost becomes salient as a potential threat. Helm would argue that if the tomato plant truly matters to me, I am rationally required to feel afraid that the frost will destroy it, angry at myself the next morning for forgetting to bring them inside, and sad that they have been killed by the frost. If, however, the frost does not arrive, I am rationally required to feel delight and a sense of relief when I wake up and my tomato plant is unharmed.

dog's. Life seemed slow and joyous and remembering those days, the girl felt that she had been on the brink of something extraordinary. She remembered the shepherd, his exuberance, energy, dignity. She remembered the shepherd and remembered being, herself, good. She had been capable of living another life then. She lived aware of happiness. (72)

When the fiancé entered her life, however, it seems that he put a lot of pressure on her to spend time with him. And the girl heeded the pressure. Moreover, because he didn't care about her dog, he put her in a situation of having to choose between looking after her Shepherd and spending time with him. This is a classic example of what Marilyn Frye (1983) refers to as a double bind. A double bind, according to Frye, is a situation where one faces a restrictive set of choices, all of which lead to a bad outcome for the agent. Say, for example, that I am habitually harassed at work by my boss. The next time this happens, I can (1) confront him (as the angry and anxious women discussed above does); (2) report his behaviour to a supervisor; or (3) do nothing. If I say nothing, the behaviour is bound to continue. If I confront my boss, he may gaslight me by insisting that I am unreasonable and misinterpreting his intentions. In the worst-case scenario, he finds a way to fire me, or makes my life at work very difficult. And finally, if I report the incident, I may face misogynistic backlash. I may not be believed; or, if I am believed, my contract may not be renewed next year, as I may be perceived as a troublemaker who could potentially contribute to a bad working atmosphere.

Similarly, we can say that the woman is in a double bind, as she cares both about her dog and her fiancé, and prioritising one over the other will mean losing one or the other. In her case, she prioritised her relationship with her fiancé, with the consequence that her dog developed behavioural problems and died tragically one night while she was out with her fiancé and his friends. Her relationship with her fiancé thereby prevented her from pursuing one of the things that she most cared about: taking care of her dog. Over the course of the story, she starts to see that her fiancé's entitlement and his demands for attention indirectly contributed to the bad situation. Instead of recognizing the dog as an intelligent, sentient being in need of care and attention, for instance, Chester perceived the dog as an annoying obstacle to having full emotional access to his fiancé, and he did everything in his power to arrange his life so as to minimise contact with the dog, including buying another house. Her life completely revolved around Chester—they socialised with his friends, and spent most of the time, it seems, at his house, as if she had no life apart from him. But—and this is key—losing her beloved Shepherd put things in perspective and initiated an epistemic shift:

> The sky was red, the water a dull silver. "I can't bear to see the Tynans again," the girl said. "I can't bear to go to another restaurant and see the sneeze guard over the salad bar."

> "Don't scream at me, darling. Doesn't any of that stuff you take ever calm you down? I'm not the dog that you can scream at."

> "What?" the girl said.

> Chester sat down on the glider. He put his hand on her knee. "I love you," he said. "I think you're wonderful, but I think a little self-knowledge, a little *realism* is in order here. You would stand and scream at that dog, darling."

> The girl looked at his hand, patting her knee. It seemed an impossibly large, ruddy hand.

"I wasn't screaming," she said. The dog had a famous trick. The girl would ask, "Do you love me?" and he would leap up, all fours, into her arms. Everyone had been amazed.

"The night it happened, you looked at the screen and you said you'd kill him when he got back."

The girl stared at the hand stroking and rubbing her knee. She felt numb. "I never said that."

"It was a justifiable annoyance, darling. You must have repaired that screen half a dozen times. He was becoming a discipline problem. He was adopting ways that made people feel uncomfortable."

"Uncomfortable?" the girl said. She stood up. The hand dropped away.

"We cannot change any of this," Chester said. "God knows if it were in my power, I would. I would do anything for you."

"You didn't stay with me that night, you didn't lie down beside me!" The girl walked in small troubled circles around the room.

"I stayed for hours, darling. But nobody could sleep on that bed. The sheets were always sandy and covered with dog hairs. That's why I bought a house, for the beds." Chester smiled and reached out to her. She turned and walked through the house, opening the door, tripping the buzzer. "Oh you've got to stop this!" Chester shouted.

(74-5)

Several aspects of the exchange are worthy of attention. First, Chester's lack of empathy is inappropriate given how distressed the woman is. Even if everything he said is true—that she would scream at the dog, and that ultimately it was her fault that he developed separation anxiety—this is not the right time to bring it up. Rather than offering *real* support, he is condescending and dismissive of her feelings, insisting that she "stop this." Moreover, the woman's somatic experience and demeanour is notable: not only does she feel numb and detached when he places his hand on her knee, but she stands up and puts distance between herself and Chester, as if repelled by his touch. She then walks anxiously, in "small, troubled circles," as if trying to make sense of the situation. Then, when he claims he would do anything to change the situation if he could, she bursts in anger, seemingly aware of his lack of empathy, which she implicitly contrasts with her dog's gestures of concern and affection in similar types of situations in the past. Whereas Chester did not lie down beside her because the bed was covered in dog hair, the Shepherd would have been by her side supporting her through this difficult time. As if feeling confident about the aptness of the anger, the scene ends with her exiting the house and leaving Chester.

What happens next is up for interpretation: does she then imagine the death of her shepherd leaping on all fours over the edge of the cliff? Does she dream about falling over the cliff? Does she commit suicide? Ultimately, it really doesn't matter how we interpret the ending of the story, however. What matters is that we understand the relationship dynamic between the protagonist and Chester as toxic and unhealthy, and that we appreciate the complexity of the events that led to the accident. Although it is fair to conclude that the woman bears some of the blame, Chester's actions and words, and her response to his callous lack of concern, suggest

that Chester might be deflecting his share of the blame onto her rather than taking some of the responsibility. But even more significant are the other clues scattered throughout the story about the degree to which he pressured her to neglect her dog. Perhaps he did so unaware that a dog needs company and stimulation and shouldn't be left alone for hours on end; more likely, however, he did not care. A classic entitled man, he doesn't need to care because all that matters to him is what's most convenient for him, even if it means calling into question his fiancé's credibility by gaslighting her.

Walking away is therefore a powerful moment in the story, since it signals to readers her awareness—even if it only manifests as a fleeting feeling—of her epistemic oppression. She comes to trust herself enough to walk away, even if she doesn't yet have the conceptual tools with which to put this new understanding into words. Readers thereby come to appreciate that she is trapped in a bird cage, and that she is starting to see the bars, to invoke Fry's metaphor of oppression. Moreover, it is also by reading about the woman's experience that Williams's readers may come away with a sense that they are not the only ones who are trapped in unhealthy relationships that compromise their epistemic agency. Williams' story is intended to be a form of consciousness raising. Finally, and most significantly, is the woman's not only leaving Chester's house, but also going to her own house after the heated exchange. Chester's house can be interpreted as symbolising the unlevel playing field on which the protagonist must recover whatever epistemic agency she has. But it is in her house where she finally begins to process the death of her Shepherd, free of Chester—a new playing field where she might come to understand the events that led to the Shepherd's death more fully or even in a new light. It is on this new playing field that we readers come to realise that the odds were always in favour of Chester and against the woman, and that even after the tragic events for which they both share the blame, Chester seems unharmed and even a little too self-confident about his blamelessness—so much so that the only thing that matters to him in that moment is that they are going to be late for dinner. The epistemic landscape, in other words, is skewed such that Chester is protected from having to examine his own culpability. By contrast, the woman struggles not only to foreground her apt anger, which crystallises through a contrast between what her dog would have done—lie by her side when she is sad or distressed—and what Chester did—leave her alone because the bed is covered in dog hair.

According to Michael Cholbi, "Grief is a path towards *freedom*." But, he adds that "the freedom in question is not the sociopolitical freedom associated with political movements" (Cholbi 2022: 187). One consequence of Williams's story, contrary to Cholbi, is that grief *is* as an important affective resource for accomplishing sociopolitical freedom. This is because, although it is true that grief initiates processes of inquiry which lead to self-knowledge, self-knowledge can only be achieved on a level epistemic playing field where individuals are taken seriously as epistemic agents and have the affective resources to exercise epistemic autonomy and narrate their own lives. As Sarah Ahmed (2017) writes, the home is political, and it is precisely because it is political that Williams transports readers from Chester's to the woman's home, where in hindsight we can reflect on Chester's perceptions of himself as an epistemic authority and on the woman's perception of herself as lacking credibility and not trusting her feelings of searching. Whenever I read this story, I am always struck by how easy it is to side with Chester. This is because we supposedly do not see things from his perspective. What I always conclude, however, is that we do not need to, because the protagonist's perspective is completely immersed in his.

Although reading the story can become an unconscious exercise in defending the epistemic landscape from critical reflection, when we feel ourselves siding with Chester, we are afforded the opportunity to learn about the *resilience* of the unlevel playing field. This is because the playing field is protected by the *affective*

practices of the community of knowers—affective habits that reinforce standing exclusion, which, as discussed above, involves calling into question the credibility, status, or *standing* of oppressed epistemic agents as knowledge producers. The reinforcement mechanism built into the epistemic landscape makes it so that it becomes resistant to critical reflection. Chester's actions gain traction and become salient within the context of practices that automatically confer authority onto men and give them the affective tools to realise their epistemic agency. These practices make it so that Chester can deflect moral responsibility onto his fiancé, which has the effect of reinforcing her uncertainty about the aptness of her anger, indicated by the way she walks in small, troubled circles. The protagonist's anger, however, registers awareness that Chester is not completely blameless. Her anger comes to the foreground when she realises that some of the blame for the dog's death falls on Chester. This too is indicated by the small, troubled circles, insofar as she is anxious to make sense of her feeling uncertain about whether she is fully to blame or whether Chester is also to blame. In other words, the protagonist is anxious to *understand* why her dog died and who is to blame, and appears to settle on an interpretation when she leaves Chester's house. She does not need her feelings of confidence to be confirmed by an admission of guilt from Chester, which she likely will never get.

Appropriate blame attribution can only happen when the epistemic landscape is not skewed to benefit dominant groups. This is because, before we assign blame and moral responsibility to ourselves and others, we must be able to use epistemic resources to *understand/interpret* the situation correctly, and to regard ourselves as epistemic authorities. If the epistemic landscape is skewed, the attribution of moral blameworthiness will not be apt. For example, Chester is certain he is blameless, but, rather than reflecting on the degree to which he put his fiancé in a double bind situation, participates in a blame attribution practice that shields him from holding himself accountable, not merely for the death of the dog, but also for enforcing epistemic norms that oppress women. Chester's epistemic landscape shields him from critical reflection, as critical reflection would reveal to him that he is benefitting from a system that accords him "unearned advantages" (Bailey 1998). The protagonist also, inevitably, participates in the same blame attribution practice, which is why she feels overwhelming guilt and shame. In her case, however, grief crystallises for her that her dog mattered to her deeply, and enables her to trust the feeling that she is not the only one to blame. It is the small, troubled circles prompted by grief, then, that expose the edges of the unlevel playing field and unleash the potential for new ways of knowing.

5. Conclusion: Resisting Epistemic Oppression

So far, we have looked within for traces of the biases that are encoded in the epistemic landscape and that inform our interpretations of the evidence and shape our emotional responses, which in the case of the protagonist of the story constitute knowledge about why her dog died and who is to blame. Her feeling of certainty indicates that she has arrived at a more accurate understanding of her situation, which is reflected in her foregrounded (and apt) anger. Anger, in turn, motivates her to leave her fiancé's house, and possibly to end her relationship with him. We have also seen that attending to the grieving, aching body can be a path towards sociopolitical freedom, as it focuses our attention on that which really matters to us. The process of inquiry it catalyses also enables us to respond to wrongdoing by attributing blame and holding others accountable for moral wrongdoing. This is because losing something that we care deeply about disrupts not only our practical identities but also our habitual ways of interacting with the epistemic landscape. Grief disrupts the workings of the affective mechanisms of the unlevel playing field, which typically function to reinforce and protect the epistemic landscape from critical reflection. Finally, we have seen the important epistemic work anxiety plays

in helping us arrive at feeling certain about how to understand a situation correctly and how to emotionally respond to it in a way that is fitting to the facts.

Although Williams's story functions as a resource for helping us to cope with affective exclusions, I hope that I have been successful at arguing that the story also tells us something about the nature of the epistemic landscape and what sustains and makes it resilient. The epistemic landscape is fuelled by affective practices which function as defence mechanisms against critical reflection. When grappling with epistemic inertia, what we are in fact dealing with are habitual ways of *feeling*, which all but guarantee that groups who benefit will be resistant to change. After all, no one wants to live in a state of anxiety and emotional discomfort. To fully address epistemic oppression, we must not only learn how to resist epistemic inertia, but also figure out how to disrupt the workings of the affective mechanisms of the epistemic landscape in dominant groups. Only then will we be able to shift the epistemic landscape and bring about new, liberatory ways of knowing.

References

Ahmed, S. 2017. *Living a Feminist Life*. Duke University Press.

Bailey, A. 1998. "Privilege: Expanding on Marilyn Frye's 'Oppression.'" *Journal of Social Philosophy* 29 (3): 104–19.

———. 2014. "The Unlevel Knowing Field: An Engagement with Dotson's Third-Order Epistemic Oppression." *Social Epistemology Review and Reply Collective* 3 (10): 62–68.

Cabrera, F. 2021. "Is Epistemic Anxiety an Intellectual Virtue?" *Synthese* 199: 13471–95

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

de Sousa, R. 2009. "Epistemic Feelings." *Mind and Matter* 7 (2): 139–61.

Dotson, K. 2014. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28 (2): 115–38.

Ellsworth, P. C. 1994. "William James and Emotion: Is a Century of Fame Worth a Century of Misunderstanding?" *Psychological Review* 101 (2): 222–29.

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

Frye, M. 1983. "Oppression." *The Politics of Reality*, Crossing Press: 1-16.

Goldie, P. 2004. "Emotion, Feeling, and Knowledge of the World." In *Thinking about Feeling: Contemporary Philosophers on Emotions*, edited by R. C. Solomon, 157–84. Oxford University Press.

Harding, S. 1986. *The Science Question in Feminism*. Cornell University Press.

Helm, B. W. 2009. "Emotions as Evaluative Feelings." *Emotion Review* 1 (3): 248–55.

———. 2010. *Love, Friendship, and the Self: Intimacy, Identification, and the Social Nature of Persons*. Oxford University Press.

Hookway, C. 2008. "Epistemic Immediacy, Doubt and Anxiety: On a Role for Affective States in Epistemic Evaluation." In *Epistemology and Emotions*, edited by G. Brun, U. Doguoglu, and D. Kuenzle, 51–65. Ashgate Publishing.

James, W. 1884. "What is an Emotion?" *Mind* 9 (34): 188–205.

———. (1894). Discussion: The Physical Basis of Emotion. *Psychological Review* 1(5): 516–529.

Kurth, C. 2018a. *The Anxious Mind: An Investigation into the Varieties and Virtues of Anxiety*. MIT Press.

———. 2018b. "Emotion, Deliberation, and the Skill Model of Virtuous Agency." *Mind & Language* 33 (3): 299–317.

Levy, N. 2016. "Have I Turned the Stove Off? Explaining Everyday Anxiety." *Philosophers* 16: 1–10.

Maibom, H. 2005. "Moral Unreason: The Case of Psychopathy." *Mind & Language* 20: 237–57.

Manne, K. 2017. *Down Girl: The Logic of Misogyny*. Oxford University Press.

Nagel, J. 2010. "Epistemic Anxiety and Adaptive Invariantism." *Philosophical Perspectives* 24: 407–35.

Peirce, Charles S. 1931–35: *The Collected Papers of Charles Sanders Peirce*, vols. 1–6, (ed.) Charles Hartshorne and Paul Weiss. Harvard University Press.

Rogers, T. 2021. "Resisting Epistemic Oppression." *HUMANA.MENTE Journal of Philosophical Studies* 14 (39): 175–93.

Roskies, A. 2003. "Are Ethical Judgments Intrinsically Motivational? Lessons from 'Acquired Sociopathy.'" *Philosophical Psychology* 16: 51–66.

Spinella, J. 2020. "A Century of Misunderstanding? William James' Emotion Theory." *William James Studies* 16 (1): 1–25.

Toole, B. 2019. "From Standpoint Epistemology to Epistemic Oppression." *Hypatia* 34 (4): 598–618.

Trajkovski, M. and T. Williamson. 2021. "Abduction, Perception, Emotion, Feeling: Body Maps and Pattern Recognition." *Philosophical Perspectives* 35 (1): 404–18.

Vazard, J. 2021. "(Un)reasonable Doubt as Affective Experience: Obsessive-Compulsive Disorder, Epistemic Anxiety and the Feeling of Uncertainty." *Synthese* 198: 6917–34.

Williams, J. 2015. *The Visiting Privilege: New and Collected Short Stories*. Vintage.

Meta-Emotions in the Context of Affective Empathy

Katharina Anna Sodoma - Ludwig-Maximilians-Universität München

Abstract

Meta-emotions are emotions about emotions, such as, for example, shame about anger. An important subset of meta-emotions constitutes a special case of co-experienced emotions, in which one emotion is directed at another emotion experienced by a subject at the same time. By making us reflectively aware of our own first-order emotions and suggesting an evaluation of them, meta-emotions enable reflection on our own emotional sensibility. In this paper, I explore the roles of meta-emotions in the context of affective empathy. I show that, whereas meta-emotions on the side of the target person increase the complexity of empathy's success conditions, meta-emotions on the side of the empathiser expand the role of meta-emotions for self-reflection and give rise to parallel phenomena of other-directed reflection. Exploring the roles of meta-emotions in the context of affective empathy allows us to better understand the significance of meta-emotions, as well as the functions of affective empathy.

Keywords: emotions, meta-emotions, empathy, affective empathy, endorsement, reflexivity

1. Introduction

"The only thing we have to fear is fear itself," said President Franklin Delano Roosevelt during his 1933 inaugural address. "And I hate how much I love you, boy," sings Rihanna. These quotes indicate a phenomenon that is familiar from everyday experience: in addition to a range of other kinds of objects, our emotions can be directed at each other. We can, for instance, feel ashamed about our anger, or relieved by our joy. These cases constitute instances of "meta-emotions," that is, emotions about emotions. Another phenomenon that is familiar from everyday life is that emotions can be "co-experienced," that is, we can experience different (and sometimes even conflicting) emotions at the same time. Although not all meta-emotions are co-experienced with their target emotions, meta-emotions about one's own emotions that occur at the same time as their target emotions constitute a special case of co-experienced emotions, in which one emotion is directed at another emotion.

Despite a recent increase in philosophical interest in emotions, meta-emotions, as well as co-experienced emotions more generally, have not yet received much attention.[1] In this paper, I discuss meta-emotions as a special kind of co-experienced emotions. My aim is to contribute to a better understanding of (co-experienced) meta-emotions by investigating the roles of meta-emotions that occur in the context of affective empathy. While some authors have pointed to potentially interesting connections between meta-emotions and empathy (Jäger and Bartsch 2006: 195, n. 20; Mendonça 2013: 393), they have left these connections unexplored. My core thesis is that investigating the roles of meta-emotions in the context of affective empathy will lead to a better understanding of meta-emotions as well as affective empathy. On the one hand, it will deepen our grasp of the significance of meta-emotions for the complexity of our mental lives, as well as for how we reflect on ourselves and others. On the other hand, it helps us better understand the mechanisms underlying some of the epistemic and interpersonal functions of affective empathy.

I begin by characterising meta-emotions and providing an overview of recent discussions (Section 2). Next, I highlight the role of meta-emotions in reflecting on our own emotional sensibility (Section 3). After introducing the notion of affective empathy (Section 4), I then bring out several implications that meta-emotions have for thinking about affective empathy. I show that meta-emotions on the side of the target person add complexity to empathy's success conditions (Section 5). If a target person experiences not only anger, but also shame about their anger, empathy will only be successful if it is responsive to both of these emotions. Subsequently, I show that meta-emotions on the side of the empathiser expand the role of meta-emotions for self-assessment (Section 6) and give rise to parallel phenomena of other-directed reflection and evaluation (Section 7). For example, if, in empathising with a target person's anger, an empathiser experiences a meta-emotion directed at this anger, this may constitute an assessment of the target person's emotion, or their own empathic reaction. In addition to expanding their role in self-reflection, and giving rise to parallel phenomena of other-directed reflection, meta-emotions that occur in the context of episodes of affective empathy figure in some of the mechanisms underlying important epistemic and interpersonal functions of affective empathy. Meta-emotions thus constitute an interesting special case of co-experienced emotions that accounts for part of the complexity of our emotional experience, and which plays an important role in our relationships to ourselves and others. Exploring the roles of meta-emotions in the context of affective empathy allows us to better understand the significance of meta-emotions, as well as the functions of affective empathy.

---

1   Recent philosophical contributions that are devoted entirely to the topic of meta-emotions include Jäger and Bartsch 2006; Mendonça 2013; Jäger and Bänninger-Huber 2015; Howard 2017; Mitchell 2020. Authors writing about emotions in general also sometimes note the possibility of emotions being directed at other emotions and briefly discuss potential implications of this case (see e.g. Deonna and Teroni 2012: 18). The term "meta-emotion" (as well as "meta-mood") is also used in psychological research. However, in this context it sometimes refers to a broader set of attitudes towards emotions (of oneself and others)—what John Gottman, Lynn Fainsilber Katz, and Carole Hooven call a "meta-emotion structure" (Gottman, Katz, and Hooven 1997)—rather than just to emotions about (one's own) emotions. Of course, meta-emotions more narrowly construed will normally be in part determined by one's overall attitudes about emotions.

2. Meta-Emotions

Meta-emotions are emotions about emotions. Imagine that, after having witnessed the embarrassing failure of a rival, you experience the pleasure of schadenfreude. However, while you are still attending to the delectable aspects of your rival's mishap, you begin to notice the nagging feeling of guilt, which shifts your attention to your schadenfreude. You are experiencing a meta-emotion: guilt about schadenfreude.[2]

That emotions can be about other emotions is a consequence of the fact that they are intentional mental states that can be directed at different kinds of objects.[3] Moreover, their intentionality is special in that they are not ways to apprehend their objects directly. Rather, they depend on other mental states, so-called "cognitive bases," such as beliefs, perceptions, or imaginings, to provide information about their objects (Deonna and Teroni 2012). In addition, emotions have a characteristic phenomenology, present their objects in an evaluative light, and have effects on attention and motivation. From these observations we can derive that meta-emotions have cognitive bases and phenomenology, they suggest an evaluation of their target emotions, and are connected to characteristic patterns of attention and action-readiness.[4]

The characterisation of meta-emotions as "emotions about emotions" allows for a distinction between intra- and interpersonal cases of meta-emotion. In the intrapersonal case, I experience an emotion about my own emotion (such as anger about my own anger), while in the interpersonal case I experience an emotion about another person's emotion (such as anger about your anger). Most authors who have written about meta-emotion explicitly restrict their discussion to the intrapersonal case.[5] In what follows, I will mainly focus on the case of intrapersonal meta-emotions as well. However, below, I will introduce a special case of meta-emotions, which arises in the context of affective empathy, and which straddles the boundary between intra- and interpersonal meta-emotion (Section 7).

Meta-emotions can be either "synchronous" or "asynchronous," that is, they can either be about emotions experienced at the same time or about emotions experienced at a prior time. For example, I can be ashamed about the fear I am currently experiencing or about fear that I experienced in the past. This is not the same as

---

2 While iterations of meta-emotions beyond the second order are possible— we can, for example, feel content with our guilt about our schadenfreude—they are likely to be limited by cognitive capacity. Like most authors, I will mainly focus on second-order emotions. Because not every emotion is accompanied by a meta-emotion, no problem of regress ensues.

3 In addition to other emotions, emotions can also be about other affective states such as moods. I can, for example, feel guilty about the cheerful mood I find myself in despite the hardship of others. However, because moods are normally not conceived of as intentional mental states that can be directed at specific objects, moods cannot be about emotions. While some philosophers think of moods as being directed at a very general object—the whole world, or one's entire environment—others take them not to be intentional at all. The notion "meta-mood" also plays a role in psychology, where it refers to different kinds of mental states directed at one's mood (Mayer and Gaschke 1988).

4 For the purposes of this paper, I will not commit to a specific theory of the emotions, but rather rely on a general characterisation of emotions that is compatible with many accounts. Philosophers working on meta-emotions from the perspective of a specific theory about emotions have espoused appraisal theories (Jäger and Bartsch 2006; see also Jäger and Bänninger-Huber 2015) and perceptualism (Mitchell 2020). Jonathan Mitchell argues that feeling and judgement theories encounter problems when applied to meta-emotions. Because, according to feeling theories, emotions are perceptions of bodily changes, they are indeed at odds with the idea that emotions can be about other emotions (Mitchell 2020: 508). Insofar as being able to account for meta-emotions is a criterion for theories of emotion (Mendonça 2013), this might present a problem for feeling theories. Mitchell's arguments against judgement theories, however, do not depend on anything specific to meta-emotions (Mitchell 2020: 511–2). They thus do not provide reason to think that judgment theories struggle to account for meta-emotions in particular.

5 Mitchell is an exception to this rule; he distinguishes between "reflexive" and "nonreflexive" meta-emotions and aims to provide a unified account for both kinds (Mitchell 2020).

saying that meta-emotions can involve occurrent or dispositional emotions. One can distinguish dispositional emotions (for example, the love for another person over many decades that involves a disposition to feel a range of emotions in different contexts and at different times) from occurrent emotions (for example, the feeling of sadness experienced at a moment of parting). Meta-emotions (as well as their target emotions) can be either occurrent or dispositional. In addition to feeling currently ashamed about my occurrent anger, I can, for example, be angry at my friend (in a dispositional way) and ashamed of this anger (in a dispositional way).

Co-experienced emotions are, by definition, experienced simultaneously by a subject. Therefore, only intrapersonal, synchronous, occurrent meta-emotions (only a subset of meta-emotions overall) constitute a special case of co-experienced emotions. In the case of co-experienced meta-emotions, the cognitive base of a meta-emotion is awareness of one's own occurrent first-order emotion.[6] In what follows, I will mainly be concerned with meta-emotions of this kind.

Emotions can be positive or negative, both in the sense of hedonic valence (experiencing them can feel pleasant or unpleasant) and in the sense of their content (whether they present their object in a positive or negative evaluative light). Experiencing fear, for example, is unpleasant, and, when we are in the grip of fear, its object appears dangerous to us. In some cases of meta-emotions, first- and second-order emotions are of the same type. We can fear fear, be angry about anger, be happy about our happiness or sad about our sadness. In these cases, the valences of the respective first- and second-order emotions are the same. In cases where first- and second-order emotions are of distinct types, we can further distinguish *valence-congruent* and *valence-incongruent* combinations. We can, for instance, feel guilty about our jealousy or happy about our love (in which case the valences of first- and second-order emotions are congruent, either both negative or both positive). But we can also be ashamed about our joy or content about grief (in which case the valences of first- and second-order emotions are incongruent, either negative-positive, or positive-negative).

Part of the appeal of reflecting on meta-emotions is that it helps us to understand some of the complexity of our emotional experience. In particular, *valence-incongruent* meta-emotions account for complex overall affective experiences that are otherwise difficult to understand. For example, Christoph Jäger and Anne Bartsch explain *survivor guilt* as a meta-emotion of guilt, directed at one's own sense of relief or gratitude about having been spared, which is at the same time experienced as inappropriate in light of the catastrophe (Jäger and Bartsch 2006: 196–9).[7] In addition, Jäger and Eva Bänninger-Huber understand *righteous indignation* in terms of a meta-

---

6   Scott Alexander Howard argues that (what I am calling) co-experienced meta-emotions constitute a counterexample to the general rule that emotions depend on cognitive bases (Howard 2017: 408–13). Howard's argument depends on the controversial assumption that a cognitive base must be a separate mental state and that awareness of one's own occurrent emotion does not constitute a separate mental state but is somehow part of the first-order emotion itself. In what follows, I will not rely on this assumption. Howard discusses the intentionality of co-experienced meta-emotions as a reason to consider meta-emotions as a distinctive kind of emotions. Another potential reason concerns their phenomenology. Meta-emotions are said to "colour" their target emotions, in the sense that the presence of a meta-emotion affects the phenomenal quality of their target emotion (Mendonça 2013, 392). In contrast to emotions about other types of intentional object, they therefore have a causal effect on their object; they change it. However, because the presence of a meta-emotion will change the overall affective state of a person, it seems difficult to ascertain whether it in fact changes the phenomenology of the target emotion or just adds complexity to the overall affective experience. I thus do not take either of the arguments for thinking that meta-emotions are distinctive that I have considered—regarding their intentional structure or phenomenology—to be conclusive. Nevertheless, I hope that the following discussion shows that the distinctive roles that meta-emotions play in reflecting on ourselves and others make them a valid object of study.

7   Not all cases of survivor guilt must be explained in terms of a meta-emotion. Perhaps survivor guilt can also be directed at first-order attitudes that are not emotions, such as beliefs about the event, as an anonymous reviewer has pointed out. Rather than defending the specific account of survivor guilt offered, I am discussing it to illustrate the potential of considering meta-emotions to elucidate some complex affective phenomena.

emotion of pride or contentment a subject feels about their own indignation (Jäger and Bänninger-Huber 2015: 790; see also Jäger and Bartsch 2006: 180–1).[8]

3. The Role of Meta-Emotions for Self-Reflection

Meta-emotions can thus constitute a special case of co-experienced emotions that helps to explain a range of complex affective phenomena.[9] In this section, I will highlight the role of meta-emotions in reflecting on our emotional sensibility.

Part of the complexity of our mental lives is due to the fact that, in addition to phenomena outside the self, our mental states can be directed at each other. That is, in addition to beliefs, desires, and emotions about objects in the world, I can, for instance, have desires about my emotions (such as the desire that an emotion persist or subside) or beliefs about my desires (such as the belief that my desire is morally problematic or laudable). This gives rise to "higher-order mental states," both in a narrow sense (mental states directed at tokens of the same type of mental state, such as beliefs about beliefs) and in a broad sense (mental states directed at other mental states of any type, such as beliefs about desires). Meta-emotions are part of this range of higher-order mental states. In particular, they are part of a range of "meta-emotional" attitudes, including other types of mental states, such as beliefs, desires, and intentions, which are directed at our emotions.

Higher-order mental states are a prerequisite for reflexivity, self-assessment, and self-regulation. They enable us to become aware of our mental states, distance ourselves from them to reflect on their status, and—where possible—adjust them in accordance with our assessment. These capacities are often taken to be an important part of what it means to be a person. A higher-order mental state that has been ascribed a particularly important role in this respect is Harry Frankfurt's notion of a second-order desire (Frankfurt 1971). Second-order desires are desires about desires, such as a desire to want to help another person or a desire not to want to take a drug.[10] As Jäger and Bänninger-Huber argue, meta-emotions play an equally crucial role in self-assessment and psychic harmony as second-order desires (Jäger and Bänninger-Huber 2015).

The state in which I have a positive higher-order mental state about a first-order mental state is often characterised as "(self-)endorsement." Endorsement thus refers to a specific constellation of mental states

8   Jäger and Bänninger-Huber also rely on the notion of meta-emotions to explain the data collected in a study that they conducted, regarding facial expressions of subjects during therapy sessions and diagnostic interviews. They found that, when discussing past emotions, subjects often showed a facial expression that did not match the emotion discussed, but which was rather expressive of contempt. Jäger and Bänninger-Huber explain this in terms of a meta-emotion directed at emotions experienced in the past (Jäger and Bänninger-Huber 2015: 795–805). Whereas their explanation of righteous indignation appeals to co-experienced meta-emotions, this study can only appeal to asynchronous meta-emotions. If both meta-emotion and target emotion were occurrent at the same time, we should expect this to bear on facial expressions. However, in the experimental setup, facial expressions that cannot be correlated with a unique emotion in the FACS (facial action coding system) were discarded. This raises some general questions about whether co-experienced emotions can be studied by relying on FACS.

9   Attending to the possibility of meta-emotions also provides a vantage point from which to address the puzzling phenomenon of aesthetic preferences for tragedy and horror stories, which elicit negative emotions. The general idea is that these fictions not only provide an opportunity for negative first-order reactions but also for positive meta-reactions. I will discuss one specific proposal along these lines below (Section 6). A similar approach is also adopted in media psychology (Bartsch et al. 2008).

10 For Frankfurt, a specific kind of second-order desire, what he calls a "second-order volition," that is, a desire for one of one's desires to determine one's will, makes freedom of the will possible. However, independently of how to understand freedom of the will, second-order desires are important for psychic harmony and self-evaluation on any kind of account of the person that takes the capacity to distance oneself from one's own attitudes and reflect on them to be essential for personhood.

that plays a role in many theories of agency (see Schroeter 2004; List and Rabinowicz 2014). It constitutes a state of psychic harmony, in which I make some of my first-order mental states truly my own by giving them my approval. The state in which I have a negative attitude towards my own first-order mental state is often described as "critical (self-)distance." Critical self-distance is an important pre-condition of self-governance. Insofar as I cannot immediately make adjustments according to my own assessment and remain committed to the first-order mental state, critical self-distance constitutes a state of internal conflict, in which I become alienated from some of my own first-order mental states.

Meta-emotions give rise to specific affective versions of self-endorsement and critical self-distance. The state in which I have a negative meta-emotion about one of my emotions constitutes "affective critical self-distance." The state in which I have a positive meta-emotion about one of my own emotions constitutes "affective self-endorsement." Jäger and Bänninger-Huber distinguish strong and moderate versions of affective self-endorsement. An agent enjoys "*moderate* affective self-endorsement" as long as they do not experience a negative meta-emotion about their first-order emotion; they enjoy "*strong* affective self-endorsement" if they experience a positive meta-emotion about this emotion (Jäger and Bänninger-Huber 2015: 808).

Emotions play an important role in our lives; they guide our attention, motivate us, and suggest an evaluation of their objects. Sometimes, emotions can play these roles without us noticing them, such as when emotions attract and repulse us in ways that are beneficial for us or when curiosity and the aesthetic pleasure a hypothesis elicits lead to scientific discovery. However, for emotions to realise their full potential, it is often crucial that we become reflectively aware of them. Meta-emotions, and meta-emotional attitudes more generally, are what enables us to do so. They are a prerequisite for reflective awareness, critical reflection, and regulation of our emotions. While we will often not be able to change our own first-order emotional reactions at will, we have more indirect strategies available. If I am alienated from my own emotional responses, I can attempt to regulate my emotional response by focusing on certain features of my situation while ignoring others, reframing the situation, or repressing the emotion. In the long run, regulating our emotions and reflecting on why they occur gives us partial control over the development of our emotional sensibility, that is, our tendency to react emotionally in characteristic ways.[11]

Meta-emotions are a special kind of meta-emotional attitude. Like other meta-emotional attitudes, they enable us to reflect on a specific aspect of our personality: our emotional sensibility. However, they contribute to reflection and self-assessment in a distinctive way. Emotions present their objects in an evaluative light. Accordingly, meta-emotions present their target emotions in an evaluative light. As Jonathan Mitchell points out, the evaluation suggested by a meta-emotion often relates to whether the target emotion is taken to be appropriate (Mitchell 2020: 515–7).[12] For example, embarrassment about anger presents anger as embarrassing. This is closely connected to whether it is appropriate to feel anger in a given situation.

---

11 The point of trying to develop our emotional sensibility is not only to suppress and weed out unwanted first-order emotions (and affirm accepted reactions), but also to cultivate new emotional reactions. In general, emotions can be directed at the absence of emotions. For instance, I can be proud of not experiencing jealousy or ashamed at not feeling saddened by a natural disaster. While these reactions do not qualify as meta-emotions narrowly understood, they are part of a wider range of attitudes towards my emotional reactions and can enable and motivate us to develop our sensibilities in this respect.

12 Emotions can be appropriate or inappropriate in different respects. We can distinguish whether an emotion is appropriate, in the sense that the object has the evaluative property it suggests, from whether it is morally or prudentially appropriate to feel it (D'Arms and Jacobson 2000). Whether these different senses of appropriateness can come apart is an open question that need not be settled here; we can allow for subjects to consider different senses of appropriateness in their assessment of their own emotions. Mitchell suggests a specific perceptualist account of how appropriateness figures in the content of meta-emotions (Mitchell 2020: 516), which I do not here subscribe to.

By suggesting an evaluation of our emotion, and guiding our attention to features of the situation (Brady 2013), meta-emotions can enable insight into the status of our first-order emotion. However, to gain insight into the status of our first-order emotion, we must be independently motivated to engage in self-assessment. In addition, we must rely on capacities for critical reflection that go beyond our emotional sensibility. Nevertheless, because they constitute a form of reflective awareness of our first-order emotions, meta-emotions can initiate reflection on our emotional reactions. Because they suggest an evaluation of our emotions, they provide input to this reflection. Therefore, at least in some cases, meta-emotions can contribute to insights about our first-order emotions which we would not have had without the occurrence of a meta-emotion.

While meta-emotions can thus enable insight, as Howard emphasises, meta-emotions need not be epistemically privileged over their target emotions (Howard 2017: 416). We can easily imagine a case in which an emotion is appropriate, while a meta-emotion (perhaps in the grip of societal expectations) provides a negative assessment of this target emotion. For instance, under the influence of gender stereotypes, a person might, in general, be ashamed about feeling sadness on any occasion. In this kind of case, in which a meta-emotion evaluates a first-order emotion negatively, the subject can resolve the conflict in two different ways. They can either adopt the view that their sadness is shameful, and regulate their emotions so that they no longer feel sad—in which case meta-emotions would be in the service of a form of emotion repression—or they can reflect on their meta-emotion and determine that their sadness is in fact appropriate and not shameful— in which case they might cease to feel shame. This kind of reflection on and regulation of meta-emotions might be especially difficult to achieve, given the complexity of the situation and the fact that meta-emotions are often inconspicuous. However, if successful, it can teach the subject something about their own attitudes towards emotions. While meta-emotions thus enable reflective awareness and assessment of our emotional reactions in a distinctive way, there is no guarantee that the presence of meta-emotion will lead to insight.[13]

4. Affective Empathy

Having characterised meta-emotions and discussed their significance for reflecting on our own emotional sensibility, I now turn to the main topic of the paper: examining the roles of meta-emotions in the context of affective empathy. In this section, I introduce the notion of affective empathy I will be working with.

The term "affective empathy" is used to denote a range of different but related states and processes. In what follows, I will be concerned with a particular kind of affective empathy—a range of closely related processes that consist of the following three elements. An empathiser:

- imagines being in a target person's situation from the target person's point of view ("imaginative perspective-taking");
- experiences a genuine emotional reaction to an aspect of this imagined situation ("affective matching"); and
- intends their emotion to represent the target person's own emotional reaction ("representation").[14]

---

13 In addition to being swayed by the assessment implicit in an inappropriate meta-emotion, we might misidentify our own first-order emotions, or misidentify or fail to identify a meta-emotion. In such cases, the presence of a meta-emotion might give rise to distinctive kinds of misunderstandings or contribute to confusion.

14 In emphasising the role of perspective-taking, the notion of empathy I will be working with resembles those employed by Olivia Bailey and Amy Coplan (Bailey 2018; 2022; Coplan 2011), among others, and is more restrictive than the influential notions of empathy developed by Heidi Maibom and Hannah Read (Maibom 2020; Read 2021). For an argument in favour of understanding affective empathy along these lines, see (Coplan 2011).

Affective empathy thus understood can be a fairly automatic and routine process, but it can also require effort. On the view presented, it is not a matter of a single mental state, but a process involving a complex of mental states (imaginative engagement, an emotional reaction, and an intention to represent the target person's emotion). Typically, affective empathy as described will start with imaginative perspective-taking, but this sequence is not necessary. A subject can also, for example, start feeling a certain way and only later become aware that this is how another person is feeling, relate it to this person's situation, and treat it as a representation of this person's emotion. Therefore, affective empathy denotes a range of closely connected processes rather than a single process.[15]

Affective empathy serves a range of epistemic and interpersonal functions. Empathy's main epistemic role is that it enables us to understand others. It can allow us to find out about which emotions they are currently experiencing and how it makes them feel (Stueber 2006; Boisserie-Lacroix and Inchingolo 2021). Empathy can also lead us to learn about the world. Adopting another person's emotional perspective can make new evaluative aspects of situations salient to us that we had previously been insensitive to (Bailey 2018; Betzler and Keller 2021). However, as some have stressed, by the same means, empathy can also corrupt our own perspective, leading us to have a less accurate view of the situation than we previously had (Betzler and Keller 2021; Paul 2021). What is less frequently noticed is that empathy can also enable us to learn about ourselves. By attending to whether and with how much ease or struggle we can empathise, we can gain insight into our "empathic sensibility," that is, our tendency to empathise or fail to empathise with certain others on certain occasions.[16]

Another important range of functions of empathy is interpersonal. Empathy is frequently experienced as a form of support. It can communicate *acknowledgment* of one's own emotional reactions but also *reassurance* regarding them. Acknowledgment means that one's emotions are seen and taken to be real and intelligible by others (Song 2015; Bailey 2022). Reassurance goes beyond that; it means that one is affirmed regarding the appropriateness of one's emotional reaction (see Betzler 2019). Both can be interpersonally important. For empathy to fulfil either of these functions, the empathiser must communicate with the target person, conveying that they can share their reaction, as well as whether they take it to be appropriate. This will typically only be the case where the empathiser is motivated by concern for the target person. Moreover, affective empathy can motivate us to act in accordance with others' projects and concerns (Maibom 2007). Because they are due to its affective dimension, empathy's effects on motivation might occur independently of antecedent concern. However, since the subject can regulate these effects, empathy will often only lead to actual support if attitudes of concern are present. Affective empathy can thus be understood in terms of a range of closely related processes that play many important interpersonal und epistemic roles.

---

15 It bears noting that the notion of perspective-taking involved can be fairly minimal. In less complete cases of empathy, an empathiser might not know which aspect of a situation a target person is reacting to, but still come to share their emotional reaction and relate it to their situation under a very general description. Moreover, this schema can also be extended to cover cases of apparent empathy in which a target person does not experience any emotional reaction (perhaps because they are unconscious or lack relevant information) by appealing to the target person's reaction under counterfactual circumstances.

16 I discuss this point in more detail below (Section 6).

5. Meta-Emotions and the Success Conditions of Affective Empathy

Having clarified the notion of affective empathy that I will be assuming, I can now begin to explore the role of meta-emotions that occur in the context of affective empathy. In episodes of affective empathy, meta-emotions can occur either on the side of the empathiser or on the side of the target person. In this section, I will focus on an implication of meta-emotions that occur on the side of the target person for thinking about affective empathy. This implication concerns empathy's success conditions.

As indicated by the above discussion, at least typically, an empathiser tries to recreate a target person's emotional reaction through imaginative engagement with their situation. This raises the question of how closely (and in which respects) the empathiser's empathic emotion must correspond to the target person's initial emotional reaction for empathy to be successful. In complete cases of affective empathy, in addition to having a similar phenomenology, an empathiser's emotion should also match the object of the target person's emotion. Whether they achieve either can depend on the detail and accuracy of their imaginative effort.

In general, correspondence is a matter of degree, and the degree of correspondence necessary for empathy to fulfil its function is a matter of context. In some situations, a relatively basic level of matching will suffice for understanding and will enable the empathiser to support the target person. In other cases, only a very high level of matching will do; anything less than that might, for example, be considered an affront by the target person. While wildly inaccurate empathy will fail to fulfil its functions, complete matching is unlikely ever to occur. Too wide-ranging are the differences between individuals and the nuances of emotional reactions. In addition, the fact that the empathiser knows that the situation is only imagined rather than actual for them will almost certainly temper the intensity of the emotional reaction felt.

It is often assumed that for empathy to be successful, an empathiser must at the very least experience an emotion of the same type as that of the target person (Coplan 2011). This can already be difficult to achieve, as it may require adjustment for differences between oneself and the target person, and opportunities for projection abound. We can easily overlook relevant differences between ourselves and others, illegitimately assume that they would react in the same way we (think we) would, and ascribe this reaction to them (Goldman 2011). The possibility of the presence of meta-emotion on the side of the target person intensifies this difficulty. To completely empathise with a target person who experiences a meta-emotion, an empathiser must experience both the second-order and the first-order emotion, experience the second-order emotion as directed at the first-order emotion (and the first-order emotion as directed at the object of the target person's first-order emotion), and take both emotions to represent the target person's emotions. This is a difficult achievement. However, it will often be necessary for empathy to be successful.

In general, whether a person endorses or critically distances themselves from their own emotional reaction can be important for understanding them and being able to offer support. If, for instance, a target person experiences jealousy and at the same time embarrassment about their jealousy, an empathiser who manages to recreate the jealousy but not the embarrassment will not succeed. Not only will they not arrive at an accurate enough understanding of the other person for most purposes, they will also not be in a particularly good position to offer them support either. Emphasising the perspicuity of their jealous reaction and the good reasons they have for it will likely only increase their embarrassment rather than alleviate it. The possibility of meta-emotions on the side of the target person thus increases the complexity of empathy's success conditions.

What about the case in which an empathiser only recreates a target person's meta-emotion but is aware of the first-order emotion through other means (or the other way around)? For instance, we can imagine a subject empathising with a target person's meta-emotion of guilt about feeling gratitude about having been spared in a disaster, without affectively sharing the first-order emotion of gratitude. Instead, empathic guilt is experienced as a result of contemplating the fact that the target person experienced gratitude. While this kind of case might fare better for the purposes of understanding, as well as in giving support, it constitutes a less complete form of empathy. It will not, for example, match the overall phenomenology of the target person's experience. In the experience of the target person, the phenomenology of first- and second-order emotions will intertwine. In particular, in the case of *valence-incongruent* meta-emotions, this can lead to a complicated overall affective state. Accurately sharing this complex overall affective state requires empathy with the meta-emotion as well as the target emotion.

While the presence of meta-emotion thus increases the complexity of empathy's success conditions, it is no reason for general scepticism about the possibility of affective empathy. Although cases in which an empathiser shares both meta-emotion and target emotion in the way described might be rare due to the complexity of the case, it is in principle possible. In addition, partial empathy, when combined with an awareness of the complexity of the target person's experience, can go a long way.

6. Meta-Emotions and Reflection on Our Empathic Sensibility

As noted above, in the context of episodes of affective empathy, meta-emotions can occur both on the side of the target person and on the side of the empathiser. After discussing an implication of meta-emotions on the side of the target person for thinking about affective empathy, I now turn to investigating the role of meta-emotions that occur on the side of the empathiser.

Because, in affectively empathising, I experience an emotion myself and am at the same time concerned with another person's emotion, episodes of empathy provide an opportunity for intra- as well as interpersonal meta-emotions to arise. The evaluation suggested by a meta-emotion that arises during an episode of affective empathy on the side of the empathiser can either be self-directed—it can concern the fact that one is empathising with this emotion—or other-directed—it can be a reaction to the target person's emotion. In this section, I examine the role of self-directed meta-emotions that occur in the context of affective empathy.

For simplicity, I will focus on a case in which the target person does not experience a meta-emotion. Let's assume that the target person experiences schadenfreude about the misfortune of a rival, but does not relate to this emotion through a meta-emotion either through affective self-endorsement or critical self-distance. In such a case, for empathy to be successful, an empathiser must only recreate this emotion sufficiently well and construe it as a representation of the target person's emotion. However, it is possible that the empathiser comes to experience a self-directed meta-emotion directed at their empathic schadenfreude.

The case of a self-directed meta-emotion that arises in the context of empathy constitutes a special case of a co-experienced (i.e., synchronous, occurrent, intrapersonal) meta-emotion, in which the target emotion is an empathic emotion. The meta-emotion's cognitive base is awareness of one's empathic emotion. In this case, only the first-order emotion is taken to represent the target person's emotion, while the second-order emotion

is experienced as the empathiser's own emotion.[17] Therefore, in this case, the meta-emotion is not part of the empathic process proper, but rather of the way in which an empathic emotion is processed. I will refer to meta-emotions of this type as "meta-empathic emotions."

The evaluation suggested by a meta-empathic emotion can either be negative—for instance, I can feel ashamed of my empathic schadenfreude—or positive—I can be proud of my empathic schadenfreude. Meta-empathic emotions enable awareness and critical reflection on one's empathic emotion in the same way that meta-emotions enable assessment of one's own first-order emotional reactions. By suggesting an evaluation of our empathic emotion, and by guiding our attention to features of the situation, they can enable insight into the status of our empathic emotion. However, to gain this insight, we must be independently motivated to assess our empathic sensibility and rely on capacities for critical reflection that go beyond our emotional reactions. If we come to an assessment of our empathic reaction, this can lead to attempts to revise our empathic sensibility by drawing on mechanisms of emotion regulation. As in the case of reflection on our first-personally felt emotions, it bears noting that, while meta-emotions can enable reflection and evaluation, they can also give rise to distinctive kinds of misunderstandings.

But is it plausible that empathisers react emotionally to their empathic emotions and spend time reflecting on them? What purpose could reflection on our empathic sensibility serve? I will approach these questions by considering a context that invites reflection on our own empathic reactions: engagement with fiction.

One of the complex affective phenomena that meta-emotions help understand is aesthetic preference for works of fiction that elicit negative emotions, such as tragedies and horror stories. While these kinds of fiction invoke emotions that are clearly negative—sadness or fear—the overall experience seems to offer something valuable to many. Attending to the possibility of meta-emotions provides a fruitful avenue of explanation. Susan Feagin, for example, explains the value of tragedy in terms of a "meta-response" to the "direct-response" of sadness (Feagin 1983). The direct response is empathic sadness, elicited by empathising with the fictional character.[18] The meta-response, which makes the whole experience worthwhile, is a complex positive affective state, a type of pleasure that constitutes a sense of common humanity. This solution allows Feagin not only to explain why engaging with tragedies is valuable, but also why engaging with art we take to be immoral does not provide the same kind of aesthetic pleasure. In the case of immoral art, we either do not have the relevant meta-response to our direct-response (because we disapprove of our direct response) or fail to have the direct-response in the first place (because we disapprove of the outlook presented by the work).[19]

While Feagin focuses on why we seem to enjoy sadness in engaging with tragedy, she suggests a strong discontinuity between the case of fictional and real-life tragedies. She argues that, while reflecting on our

---

17 By contrast, in the case where an empathiser tries to match a target person's meta-emotion, as discussed above, they must take both the meta-emotion and the first-order emotion as a representation of the target person's emotions.

18 Whether empathy and imagination play a role in our engagement with fiction, and if so which, is a difficult question, beyond the scope of this paper. I will assume that, at least sometimes, our first-order reactions to fictional scenarios are based on empathy with fictional characters. Feagin uses the term "sympathy" to characterise our direct response to tragedy. Because she does not elaborate on the notion of sympathy at issue, it is not entirely clear whether she has a similar mechanism in mind.

19 This connects the discussion of meta-emotions in the context of fiction to the topic of "imaginative resistance" (Gendler 2000). Whether Feagin's account is successful depends on details of the view, for instance, concerning how to understand the meta-response that constitutes a sense of common humanity. These questions are beyond the scope of this paper. Rather than defending the account in detail, I rely on Feagin's account in broad outline to motivate the idea that meta-empathic emotions can play a role in self-reflection.

own direct response is appropriate in the context of fiction—after all, we know the events not to be real—as a reaction to real tragedy, it would be a quite appalling testament to one's self-involvement. Feagin's remark has plausibility.[20] Nevertheless, while it seems right, as Feagin suggests, that, when attending to a real-life tragedy, our emotional energies should be consumed by our first-order reactions, reflecting on our empathic reactions can sometimes be appropriate. We can, for instance, imagine a person being content with their empathy with reactions to more subtle forms of insult that they were previously unable to empathise with, because this indicates a certain development of their own sensibility, which they were trying to achieve. We can also imagine a person being relieved that they can empathise with their friend's joy, because they had to overcome obstacles to be able to do so. Spontaneous empathy with reactions one finds appalling can be a proper occasion for meta-empathic feelings of shame or guilt, and inability to empathise with such reactions can be a proper occasion for meta-empathic contentment.

While the main epistemic function of empathy is that it enables us to learn about others and the world from their perspective, a less frequently discussed epistemic function of empathy is to enable us to learn about ourselves by attending to our own empathic reactions. Meta-emotions that occur during episodes of affective empathy are part of what enables us to do so. Engaging with our empathic sensibility in this way is useful because, given the epistemic and interpersonal functions of empathy, who we empathise with in which situation matters. Attending to our own empathic emotions constitutes an important form of self-assessment, which is enabled by meta-empathic emotions.

While meta-emotions that arise in the context of episodes of affective empathy on the side of the target person make successful empathy even more difficult to achieve, meta-emotions that occur on the side of the empathiser thus expand the role of meta-emotions for self-reflection. Attending to this role of meta-emotions helps explain an important epistemic function of affective empathy, to wit, how empathy enables us to learn about ourselves.

7. Meta-Emotions and Reflection on Others' Emotional Sensibility

As noted above, meta-emotions that occur on the side of the empathiser in the context of affective empathy can either be self-directed or other-directed. Having discussed the case of self-directed reflection through meta-emotions that occur in the context of affective empathy above, in this section, I examine the role of other-directed meta-emotions that occur in the context of affective empathy.

While meta-emotions that arise during an episode of affective empathy can be in the service of taking a stance on one's own empathic sensibility, they can also suggest an evaluation of the target person's emotional reaction. It is, for instance, equally possible that an empathiser experiences indignation about the target person's schadenfreude, which they empathise with. While in this case the cognitive base is, again, awareness of one's empathic emotion, it constitutes a different kind of meta-empathic emotion, in which the focus is on the target person's emotional reaction, represented by the empathic emotion. This constitutes a special case of meta-emotion, which is not easily classified with respect to the distinction between intra- and interpersonal emotions. On the one hand, the subject experiences both an empathic first-order and the second-order

---

20 In fact, in the context of a structurally similar analysis, Gregory Currie and Ian Ravenscroft suggest that, because we attend to our own response rather than the work itself, even in the context of engaging with fiction, "cultivated responses to tragedy never entirely evade the suspicion of self-congratulation" (Currie and Ravenscroft 2002: 203).

emotion, which suggests that it is a case of intrapersonal meta-emotion. On the other hand, they take their empathic first-order emotion to represent the target person's emotion, which they are concerned with, suggesting that it is a special case of interpersonal meta-emotion. The case of other-directed meta-empathic emotions thus straddles the boundary between intra- and interpersonal meta-emotions.

Other-directed meta-emotions that occur in the context of affective empathy enable an empathiser to reflect on and assess the target person's emotional reaction in the same way that self-directed meta-emotions enable them to assess their own emotions, including their empathic emotions. By suggesting an evaluation of the target person's emotion, and by guiding our attention to features of the situation, they can enable insight into the status of this emotion. However, to gain this insight, we must be independently motivated to assess the target person's emotional sensibility and to rely on capacities for critical reflection that go beyond our emotional reactions. While, in this case, there is no possibility of regulating emotions in the sense that we regulate our own emotions—for instance, by refocusing attention or reframing—we might be able to use emotion regulation strategies in a more encompassing sense. For instance, we might be able to help the target person process their own emotions by suggesting interpretations of their emotions and offering comfort and emotional support.[21] In addition to enabling reflection and evaluation, meta-emotions can contribute to misunderstandings in this case as well.

The way in which self- and other-directed assessment through meta-emotion can figure in episodes of affective empathy has interesting consequences for thinking about empathy. For one, it elucidates the connection between empathy, endorsement, and critical distance. Other-directed meta-emotions that occur in the context of affective empathy give rise to phenomena of other-directed affective endorsement and critical distance that run parallel to affective self-endorsement and critical self-distance. In the case where the empathiser experiences indignation about the target person's emotion, which they nonetheless empathise with, they critically distance themselves from the target person's emotion. Were the empathiser to experience, for example, admiration towards the target person's schadenfreude, which they empathised with, instead, this would constitute affective endorsement.

This shows that empathy can be accompanied by critical distance as well as endorsement (where both endorsement and critical distance can be implemented through a meta-emotion or another type of meta-emotional attitude). However, as Jäger and Bänninger-Huber argue for the case of affective self-endorsement, having an emotion and failing to critically distance oneself from it already constitutes a moderate form of endorsement (Jäger and Bänninger-Huber 2015: 808). Maintaining critical distance while empathising is thus cognitively more demanding than endorsing the target person's emotion. It requires an additional attitude directed at one's empathic emotion, whereas moderate endorsement is the default. This explains why, although empathy need not come with endorsement, it tends towards endorsement.

Attending to the connection between empathy and endorsement further helps explain the mechanisms underlying some of the interpersonal and epistemic functions of affective empathy. For one, it helps explain the potential for empathy to enable us to learn about the world form another person's perspective. In affectively empathising, we take on another person's emotional perspective on the object of their emotion. While it is possible to distance ourselves from this reaction, doing so is cognitively more demanding. The default is a moderate form of endorsement. There is thus at least some pressure to align our own assessment with the one implicit in the target person's emotion, which we share in empathy (Bailey 2018; Betzler and Keller 2021).

---

21 I thank an anonymous reviewer for pointing this out.

In cases where the target person's emotional reaction is appropriate, this can lead us to gain new evaluative insights. However, attending to the connection between empathy and endorsement also helps us explain how empathy can corrupt our own perspective, as some have stressed (Betzler and Keller 2021; Paul 2021). In cases where the target person's emotion is inappropriate, empathy might lead us to adopt the assessment suggested by that emotion, nonetheless. In spite of this risk, attending to the connections between empathy and endorsement helps us to understand an important epistemic function of affective empathy: learning about the world by taking up others' perspectives.

Attending to the connections between empathy and endorsement also enables us to spell out some of the mechanisms underlying an important interpersonal function of empathy. Unless the empathiser distances themselves from their empathic emotion, empathy comes with at least moderate affective endorsement. Therefore, empathy can play the important interpersonal role of reassuring a target person concerning the appropriateness of their emotional reactions.[22] Moreover, empathy that is accompanied by a positive meta-emotion and thus strong affective endorsement can fulfil the function of reassurance particularly well. While empathy that does not come with either moderate or strong endorsement cannot fulfil the role of reassurance, it can still fulfil the role of acknowledgement.

In addition to expanding the role of meta-emotions for self-reflection, meta-emotions that occur in episodes of affective empathy on the side of the empathiser thus give rise to parallel phenomena of other-directed reflection. Attending to this role of meta-emotions helps to elucidate the connection between empathy, critical distance, and endorsement. This, in turn, helps in understanding the mechanisms underlying an important epistemic function of empathy—learning about the world from the point of view of other. However, it also elucidates how taking on others' perspectives can affect our own perspective negatively. In addition, it helps to explain an important interpersonal function of empathy—providing reassurance regarding the appropriateness of one's emotional reactions. Exploring the role of meta-emotion in the context of affective empathy thus not only deepens our grasp of the role of meta-emotion for self- and other-directed reflection and evaluation; it also contributes to our understanding of affective empathy and its epistemic and interpersonal functions.

8. Conclusion

In this paper, I have discussed meta-emotions as a special case of co-experienced emotions that play an important role for reflection. My main aim was to contribute to a better understanding of (co-experienced) meta-emotions by investigating the roles of meta-emotions in the context of affective empathy. My core thesis was that investigating the roles of meta-emotions in the context of affective empathy leads to a better understanding of meta-emotions as well as affective empathy.

I have shown that, while meta-emotions on the side of the target person increase the complexity of empathy's success conditions, meta-emotions on the side of the empathiser expand the role of meta-emotions for self-reflection and give rise to parallel phenomena of other-directed reflection and evaluation. Exploring these roles of meta-emotions in the context of affective empathy thus deepens our grasp of the significance of meta-emotions for the complexity of our mental lives, as well as for reflecting on ourselves and others. At the same time, it helps us to better understand the mechanisms underlying some of the epistemic and interpersonal

---

22 Elsewhere, I discuss the importance of reassurance through empathy in contexts of emotional gaslighting (Sodoma 2022).

functions of affective empathy. Attending to the role of meta-empathic emotions in reflecting on our own empathic sensibility elucidates the way in which empathy can enable us to learn about ourselves. Attending to the role of meta-emotions in other-directed assessment through empathy helps us to better understand the connection between empathy, endorsement, and critical distance. This, in turn, helps us to understand how empathy enables us to learn about the world, as well as how it can corrupt our perspective. In addition, it allows us to clarify how and when empathy can play the important interpersonal role of reassurance. Reflecting on the role of meta-emotions in the context of affective empathy thus improves our understanding of both meta-emotion and affective empathy.

References

Bailey, O. 2018. "Empathy and Testimonial Trust." *Royal Institute of Philosophy Supplement* 84: 139–60.

———. 2022. "Empathy and the Value of Humane Understanding." *Philosophy and Phenomenological Research* 104 (1): 50–65.

Bartsch, A., P. Vorderer, R. Mangold, and R. Viehoff. 2008. "Appraisal of Emotions in Media Use: Toward a Process Model of Meta-Emotion and Emotion Regulation." *Media Psychology* 11 (1): 7–27.

Betzler, M. 2019. "The Relational Value of Empathy." *International Journal of Philosophical Studies* 27 (2): 136–61.

Betzler, M. and S. Keller. 2021. "Shared Belief and the Limits of Empathy." *Pacific Philosophical Quarterly* 102: 267–91.

Boisserie-Lacroix, C. and M. Inchingolo. 2021. "Empathy for a Reason? From Understanding Agency to Phenomenal Insight." *Synthese* 198 (8): 7097–118.

Brady, M. S. 2013. *Emotional Insight. The Epistemic Role of Emotional Experience.* Oxford University Press.

Coplan, A. 2011. "Understanding Empathy: Its Features and Effects." In *Empathy: Philosophical and Psychological Perspectives*, edited by A. Coplan and P. Goldie, 3–18. Oxford University Press.

Currie, G. and I. Ravenscroft. 2002. *Recreative Minds: Imagination in Philosophy and Psychology.* Oxford University Press.

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

Deonna, J. and F. Teroni. 2012. *The Emotions: A Philosophical Introduction.* Routledge.

Feagin, S. L. 1983. "The Pleasure of Tragedy." *American Philosophical Quarterly* 20 (1): 95–104.

Frankfurt, H. G. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1): 5–20.

Gendler, T. S. 2000. "The Puzzle of Imaginative Resistance." *The Journal of Philosophy* 97 (2): 81.

Goldman, A. I. 2011. "Two Routes to Empathy: Insights from Cognitive Neuroscience." In *Empathy: Philosophical and Psychological Perspectives*, edited by A. Coplan and P. Goldie, 31–44. Oxford University Press.

Gottman, J. M., L. F. Katz, and C. Hooven. 1997. *Meta-Emotion: How Families Communicate Emotionally.* Lawrence Erlbaum Associates.

Howard, S. A. 2017. "Metaemotional Intentionality." *Pacific Philosophical Quarterly* 98 (3): 406–28.

Jäger, C. and A. Bartsch. 2006. "Meta-Emotions." *Grazer Philosophische Studien* 73 (1): 179–204.

Jäger, C. and E. Bänninger-Huber. 2015. "Looking into Meta-Emotions." *Synthese* 192 (3): 787–811.

List, C. and W. Rabinowicz. 2014. "Two Intuitions about Free Will: Alternative Possibilities and Intentional Endorsement." *Philosophical Perspectives* 28 (1): 155–72.

Maibom, H. L. 2007. "The Presence of Others." *Philosophical Studies* 132 (2): 161–90.

———. 2020. *Empathy*. Routledge.

Mayer, J. D. and Y. N. Gaschke. 1988. "The Experience and Meta-Experience of Mood." *Journal of Personality and Social Psychology* 55 (1): 102–11.

Mendonça, D. 2013. "Emotions about Emotions." *Emotion Review* 5 (4): 390–6.

Mitchell, J. 2020. "Understanding Meta-Emotions: Prospects for a Perceptualist Account." *Canadian Journal of Philosophy* 50 (4): 505–23.

Paul, L.A. 2021. "The Paradox of Empathy." *Episteme* 18 (3): 347–66.

Read, H. 2021. "Empathy and Common Ground." *Ethical Theory and Moral Practice* 24 (2): 459–73.

Schroeter, F. 2004. "Endorsement and Autonomous Agency." *Philosophy and Phenomenological Research* 69 (3): 633–59.

Sodoma, K. A. 2022. "Emotional Gaslighting and Affective Empathy." *International Journal of Philosophical Studies* 30 (3): 320–38.

Song, Y. 2015. "How to Be a Proponent of Empathy." *Ethical Theory and Moral Practice* 18 (3): 437–51.

Stueber, K. R. 2006. *Rediscovering Empathy: Agency, Folk Psychology, and the Human Sciences*. MIT Press.

Indeterminacy in Emotion Perception: Disorientation as the Norm

Ditte Marie Munch-Jurisic - University of Virginia and University of Copenhagen

Abstract

Most psychological and philosophical theories assume that we know what we feel. This general view is often accompanied by a range of more specific claims, such as the idea that we experience one emotion at a time, and that it is possible to distinguish between emotions based on their cognition, judgment, behaviour, or physiology. One common approach is to discriminate emotions based on their motivations or ultimate goals. Some argue that empathic distress, for instance, has the potential to motivate empathic concerns; personal distress, on the other hand, is self-oriented and motivates egoistic concerns. In this paper, I argue against this and similarly teleological views of emotions and affect. Through a close study of the emotional breakdown of an American drone operator, I make the case that understanding our emotions entails much more ambiguity than dominant theories assume. In our emotional lives, disorientation and confusion are often the norm.

Keywords: basic emotions, distress, moral injury, ambiguity, indeterminacy, motivation

1. Introduction[1]

In the philosophical and psychological literature on emotions, a prevailing assumption is that we know what we feel. This general view is often accompanied by a range of more specific claims, such as the idea that we experience one emotion at a time, or at least that one emotion is always felt more strongly than another. This is, for example, a standing assumption in most thinking that aligns with some version of the basic emotion theory.[2] From this perspective, each emotion has certain defining features such that it is possible to distinguish one from another based on their associated cognition, judgement, behaviour, or physiology (Lench, Flores, and Bench 2011: 61; Kurth 2022: 61). Theorists disagree on how exactly to individuate emotions

---

2   For a recent defence of this prevalent view in the study of emotions, see Keltner et al. 2019; Scarantino and Griffiths 2011.

but, for many, these differences are not just epistemological but metaphysical and ontological facts about the world. For such thinkers, there exist distinct and clearly demarcated kinds of emotions—sometimes conceived as natural kinds—which we can track through biological, physiological markers like facial expressions, neural mechanisms, and behaviour patterns (Hutto, Robertson, and Kirchhoff 2018; Ekman 1992).

One common approach has been to distinguish emotions based on their motivations (Lench, Flores, and Bench 2011: 838; Kurth 2022: chap. 2.4). Psychologist Daniel Batson, for example, argues that the distress we feel over someone's suffering is *either* empathic distress or personal distress, vicariously induced. From his perspective, we can distinguish these two emotions based on their "ultimate goals" (Batson, Fultz, and Schoenrade 1987; Batson 2009). Empathic distress has the potential to motivate empathic concerns; personal distress is self-oriented, motivating only egoistic concerns. In this paper, I argue against this and similar teleological views of emotions and affect grounded in assumptions that rudimentary, physiological responses like distress have a specific direction or end goal.[3]

In fact, there is a significant amount of ambiguity in how we experience our emotional lives. Often, we interpret our emotional experiences through multiple explanations that may conflict and change over time. For a significant range of experiences, emotional indeterminacy is the norm. To portray emotion perception as something that happens through a single, distinct emotion concept is a theoretical construct, devised in philosophy books and experimental psychology labs, and not grounded in lived experience.

In Robert Frost's "Mending Wall" (1914), the poet explores the theme of boundaries, questioning their necessity and function. The poem's narrator and his neighbour meet every spring to mend the stone wall between their properties. While his neighbour maintains that "Good fences make good neighbors," the narrator doubts the need for the wall, and muses on the natural forces that topple the stones year after year: "Something there is that doesn't love a wall, That wants it down."

For the purposes of this paper, this something that resists our neat constructions is our affective and emotional lives. As philosophers and psychologists, we try to make sense of emotions with categories, names, and distinctions, but the nature of emotions and our affective responses continues to remain theoretically obscure (Leys 2017). Of course, we sometimes experience emotions in straightforward ways—"I'm just so happy!" we exclaim—but standard theories of emotion fail us in the many cases where ambiguity and ambivalence are present, when we are not quite sure what we feel. This is easy to demonstrate when we remove ourselves from the realm of theory and study emotions in context.

To ensure a focused discussion in service of the broader point, I make my case by centring on the feeling of distress we experience at the suffering of others. Many psychological and philosophical theories have attempted to categorise this particular affective state, but here I focus primarily on the account of experimental psychologist Daniel C. Batson and colleagues, who conceptualise this affective state within a teleological account of emotions and affect, distinguishing between emotions according to specific end goals (Section 1).

---

3   Teleological views of emotions are found across a wide spectrum of theories in philosophy and psychology. They come in different variations and do not form a uniform school of thought, but what they share is the metaphysical assumption that emotions have some form of purpose or endgoal. From this perspective, there exist distinct emotions with a core motivating feature, and that the meaning of the emotion is derived from this feature. For Andrea Scarantino, for example, each emotion involves specific action tendencies that motivate us (Scarantino 2014). Appraisal theory points to the idea of core relational themes which distinguish each emotion (Lazarus 1991). But the broader idea—that each emotion should be characterised and evaluated according to their motivational features—is widespread. See for example normative debates on what constitutes the core motivational features of anger in Silva 2021a.

I next consider an example of emotion perception through the distress and disorientation experienced by an American drone analyst, Chris Aaron (Section 2), and demonstrate the insufficiency of Batson's framework in understanding such a case. I then turn to philosophers Peter Goldie and Hannah Arendt to argue that emotion perception is a process without a determinate endpoint, and that disorientation is a fundamental, constitutive feature of human nature and ethical life, which must be taken seriously in any convincing theory of emotions (Sections 3–5). As the case of the drone analyst may in some respects be considered exceptional, I also include a discussion of more ordinary examples of disorientation in emotion perception.

To be clear, what follows is an argument dedicated to the first-person perspective, and how we come to understand our own emotions. But it is not just an epistemological argument. What emotions are (their ontology or metaphysics) cannot be separated from the way we perceive them.

Emotion perception as such is not merely characterised by *ambivalence*, that is, it is not that the emotion concepts are epistemically accessible, but that we do not know which one to choose from. Nor is emotion perception merely characterised by *ambiguity*, where we do not know what we are feeling at a given time, but that we could do so were the specific, distinct emotion categories to become available.[4] Even beyond these specific obstacles, I argue that it is often entirely indeterminate what a bodily feeling means. It is our interpretation of the feeling that gives it content, meaning, and moral motivation (Barrett 2017; 2022). In such cases, bodily feelings do not determinately reveal or mean any one thing. We make sense of our physiological distress and other emotions by the application of hermeneutic equipment availed by our environment (Munch-Jurisic 2021; 2022: chap. 4).

2. A Teleological Account of Distress[5]

Why do we feel distress when we witness or hear about the suffering of another? The question has occupied philosophers for centuries, but psychologists began to explore it in laboratory experiments during the middle of the twentieth century. Simner and colleagues, for example, found that human beings are born with the capacity to feel vicarious distress. They observed how newborns became upset and started crying when they were exposed to the crying of other infants, but not when they were exposed to a recorded version of their own cry. This early reactive crying is a primitive and innate arousal mechanism that disappears in the first half year of the infant's development. The capacity to feel distress over the suffering of others remains intact, but whose suffering the child reacts to, and how they are motivated to respond to it, will depend on the child's cognitive development and the social and moral commitments it picks up from its environment (Simner 1971; Hoffman 2000: 65).

The literature on vicarious distress is vast.[6] For the purposes of this discussion, the key point is that some psychologists conceive of vicarious distress within a teleological model of emotion. Developmental psychologist Martin Hoffman, for example, argues that the capacity to feel vicarious distress is a precursor to the capacity to feel empathic concern (Hoffman 2000). For this reason, Hoffman speaks only of "empathic distress"; for him, this affective state is meant to develop into empathic concern and has a prosocial function. From this perspective, the capacity has an inherent moral direction that may be distorted or thwarted by egoistic or antisocial concerns.

---

4   I thank Marco Viola for encouraging me to specify this point.
5   This section builds on research from chapters 2 and 3 in my book *Perpetrator Disgust: The Moral Limits of Gut Feelings* (2022).). Because of copyright restrictions, this particular section is excluded from the open access license. Thank you to Oxford University Press for the right to reproduce parts of my previous arguments.
6   For an overview of this literature, see Maibom 2014a.

Experimental psychologist Daniel Batson also proposes a teleological account of vicarious distress, though he divides the emotional landscape in a different way than Hoffman. In his view, we can differentiate between different *kinds* of distress based on their ultimate goals.[7] *Personal distress* is vicariously induced distress (i.e. triggered by others), but is fundamentally self-centred, an emotion that directs the agent's attention inwards, to oneself and one's own needs (Batson, Fultz, and Schoenrade 1987: 19–22).[8] *Empathic concern*, on the other hand, is not just triggered by others, but is directed toward the suffering of the other, and entails a prosocial concern and a motivation to alleviate the other's pain. Despite their crucial differences, both Hoffman and Batson's research relies on the teleological idea that certain affective states have an inherent direction towards a specific emotion, accompanied by a distinct motivation. For researchers like Batson, this implies that we can distinguish between emotions (and their accompanying affective states) based on characteristics that speak to their ultimate goals.

How exactly are we supposed to make such distinctions and discern different kinds of distress? To demonstrate that vicarious distress and empathic concern are two distinct emotions, Batson and his colleagues conducted a number of experiments. They exposed test subjects to a range of cases with victims in varying situations of need, such as stories of tragic car accidents or a video of a woman being exposed to increasingly painful shocks (Batson, Fultz, and Schoenrade 1987: 29). The subjects where then asked to report how they felt during the experiment through a multiple-choice questionnaire with a list of adjectives to choose from. If the subjects indicated feeling "alarmed, grieved, upset, worried, disturbed, perturbed, distressed, and troubled," Batson and his colleagues inferred they felt personal distress. If they ticked "sympathetic, moved, compassionate, tender, warm, and softhearted", the researchers inferred that they felt empathic concern. In another set of experiments, researchers tested the subjects' responses to appeals for help. The general conclusion from these experiments was that test subjects who had reported emotions classified as traits of personal distress were less likely to respond to calls for help than those who had reported emotions associated with empathy and empathic concern.

Even in these first experiments, which laid the foundation for decades of further research, an important irregularity appeared. When hearing some of these stories, subjects ticked emotion words that expressed *both* empathic concern and personal distress. To account for these irregularities, Batson and his colleagues were forced to concede that, in some need situations, personal distress and empathy are closely intertwined (Batson, Fultz, and Schoenrade 1987: 29).

This concession is crucial. Batson's experiments set out to demonstrate that there was no prosocial potential in vicariously induced feelings of personal distress, and that true empathic concern only springs from feelings of empathic distress. But the experiments demonstrated that vicariously induced distress can be simultaneously self-centred and other-concerned.[9] When people were exposed to the shocking scenarios of other people suffering, they felt a range of emotions with different and conflicting motivational states, and they would have likely reported an even wider diversity of emotion words, had Batson's experiments opted for an open-ended model of questioning.[10]

---

7   In a more recent publication Batson clarifies that his reference to ultimate goals is related primarily to motivation: "'Ultimate' refers to means–end relations, not to a metaphysical first or final cause, and not to biological function. An ultimate goal is an end in itself. In contrast, an instrumental goal is a stepping stone on the way to an ultimate goal" (Batson 2014: 43).

8   Here I rely on Batson's foundational experiments. For a more recent version of the same, see Batson 2009. It is important to note that Batson does not argue that prosocial behaviour is necessarily moral behaviour.

9   See also Maibom's discussion of this point (2014b: 7-10).

10 For examples of such open-ended experimental set-ups, see Hoemann et al. 2020; Betz, Hoemann, and Barrett 2019.

The central point is that, in concrete instances of emotional states (and not neatly isolated, abstract examples), it is inherently difficult to distinguish between different kinds of emotional states. Importantly, this is not just Batson's problem. Any theory of emotion that proposes to classify emotions into singular categories must grapple with this fact. Most philosophical and psychological theories of emotions assume that the intentionality of an emotion is clear. In other words, they assume that we can determine what a subject feels distress towards, how the subject experiences this distress (as self- or other regarding), and the action that the emotion motivates.[11] But if we assume this kind of transparency, we risk basing our interpretations on a distorted and skewed understanding of our emotional lives (Maibom 2014b: 7). In fact, as we explore in a moment, the phenomenological experience of our emotional responses is diverse and often based on conflicting emotions, affective and physiological experiences that we find hard to understand and interpret. Because of the indeterminate character of these emotional and physiological experiences, agents often do not know exactly why they feel distressed, or what their distress is directed at.

3. "My Hands Stopped Working": Making Sense of Distress

Chris Aaron joined the drone program as a 24-year-old, out of curiosity. When he first saw a Predator drone unleash its lethal payload, he found it surreal and awe-inspiring (Press 2021: 100). He recalls feeling thrill and "a surge of adrenaline" as he saw the "the camera zooming in, the laser locking on, a plume of smoke rising above the scorched terrain where the missile struck" (Press 2021: 100). Afterwards he exchanged congratulatory high fives with the other analysts in the room. From 2004 and onwards, he worked for different agencies and contractors. But one day in 2010 he experienced something strange as he mulled the terms of a new job opportunity:

> He began to fall apart physically. The distress began with headaches, night chills, joint pain, a litany of flu-like symptoms that, every few weeks, would recur. Soon, more debilitating symptoms emerged: waves of nausea, eruptions of skin welts, chronic digestive problems. Chris had always prided himself on his physical fitness. Now, suddenly, he felt frail and weak, to the point that working for the contractor was out of the question. "I could not sign the paperwork," he said. Every time he sat down to try, "my hands stopped working—I was feverish, sick, nauseous." (Press 2021: 101)

He eventually moved back to live with his parents, very unwell; his friends were barely able to recognise him—"It wasn't the same guy" (Press 2021: 114). Doctors could not specify a diagnosis, so he began experimenting with fasting, yoga, and Chinese medicine. His days were spent in a "fog of doom," and at night he had nightmares about innocent people being maimed and killed. One recurring dream was akin to the infamous aversion therapy scenes in Stanley Kubrick's *A Clockwork Orange*. Stuck in a chair, unable to avert his gaze, Aaron was forced to watch scenes of violence unfold up close: "It was as though my brain was telling me: Here are the details that you missed out on," he said. "Now watch them when you're dreaming" (Press 2021: 101–2). The hardest part for Aaron was to reckon with the feelings of joy he had experienced as a drone pilot. He had enjoyed the power that came with his position. It had been exciting. But in the years that followed, his mood darkened:

> He withdrew, sinking into a prolonged funk shadowed by shame and grief. He avoided seeing friends. He had no interest in intimate relationships. He struggled with quasi-suicidal thoughts, he told me, and with facing the depth and gravity of his wounds, a reckoning that began in earnest only in 2013

---

11 As this line of argument suggests, my criticism targets not only teleological theories of emotions, but any theory that takes for granted that the intentionality of an emotional episode is always clear and transparent. I will reserve this discussion of intentionality for another occasion.

when he made his way to the Omega Institute in Rhinebeck, New York, to attend a veterans' retreat run by a former machine gunner in Vietnam (Press 2021: 119–20).

At the veteran's retreat, Aaron finally felt capable of discussing his feelings. He was, for example, able to talk about the strategic use of "diffusion of responsibility" in the drone pilot program, which aimed to mitigate feelings of personal responsibility that individual operators felt for their actions. But Aaron *did* feel responsible and culpable, and began to interpret his feelings as a form of *moral injury* (Press 2021: 120). Like many other veterans, he found the diagnostic concept of post-traumatic stress disorder (PTSD) to be insufficient to cover the range of complex feelings and moral uneasiness that he experienced (Boudreau 2011; Litz et al. 2009).

What significance should we assign to the initial shaking that Aaron experiences in the course of his physiological breakdown? From journalist Eyal Press's narrative, it is tempting to characterise the shaking as a form of *moral distress*, maybe even a moral reckoning, especially because Press couples this with Aaron's more explicit forms of shame and guilt that only become explicit later in the process. Should we understand his body as sending a message? Is the shaking an attempt to communicate some deep moral truth, or the voice of his conscience?[12] As I will argue in the remainder of this paper, it is crucial to refrain from such conjectures. Though they may be tempting to draw, such conclusions project values on indeterminate physiological distress, offering teleological interpretations of Aaron's emotional turmoil. Instead, we must begin by recognising that Aaron was for a very long time in a state of complete disorientation about the meaning of his physiological distress. In his own recounting it was not until 2013—three years after the first episode of shaking—that he would begin to engage in a more explicit form of moral reckoning about his actions.

In the initial moments when he began to shake and feel ill, he was certainly experiencing some form of acute *conflict*, but it would be an overinterpretation to conclude at this point that his distress contained a pre-determined moral direction. In a state of conflict and confusion, Aaron was fundamentally unclear about how to account for his experience and its implications. Eventually, at the veteran's retreat, when Aaron began to interpret his initial physiological distress as a *moral* distress, his interpretation was aided by the specific context of the veteran's retreat and the novel hermeneutic tools that it made available to him. Aaron finds a form of closure in the concept of moral injury and the adjoining therapeutic practices that he is offered. But, as we explore further in the following section, it is crucial not to see Aaron's story within a teleological framework of emotion, in which the construct of moral injury is seen as the sole, inevitable interpretation of Aaron's initial physiological state—as if his final feelings of compunction, guilt, and shame were already present as seeds within his initial physiological distress.[13] Already in 2008, when he was applying for his second job, Aaron experienced a similar

---

12 For a discussion of the moral value of uncontrollable affect in similar cases, see Charlie Kurth's discussion of the physical discomfort that 18th century abolitionist John Woolman felt when he had to write a bill of sale for an enslaved woman (Kurth 2018b: 185).

13 Whether or not the construct of moral injury should be considered a diagnosis is a controversial topic among scholars and clinicians studying the phenomenon. As the inventors of the Moral Injury Scale (MIOS) put it in a recent review article: "The concept of moral injury has recently emerged in the research literature as a separate aspect of trauma exposure, distinct from posttraumatic stress disorder (PTSD). Moral injury is not classified as a mental disorder. It is a dimensional problem that can have profound effects on critical domains of emotional, psychological, behavioral, social, and spiritual functioning" (Litz and Kerig 2019: 99). For now, scholars such as Litz and Kerig argue that moral injury is not a unique diagnosis because it is not distinct enough from PTSD. They believe a slightly amended version of the DSM-5's definition of PTSD will be able to account for moral injury (through the new definition of complex PTSD). Like other scholars, I believe the phenomenology of moral injury is inadequately captured by the causal model of PTSD, and better understood as an ethical and moral struggle beyond psychological models (Boudreau 2011; Molendijk 2018; Wiinikka-Lydon 2017; 2019; 2022). Moreover, I see the current conceptualization of moral injury as a time-specific, hermeneutic equipment with its own inherent normativity, which is comparable, albeit not similar, to constructs like "partisan hysteria" from other contexts (Antić 2016: chap. 5).

physiological disruption, which he did not interpret through a moral lens. During a lie detector test, he lost circulation in his arm. Feeling hectored by the questions, he abruptly left the examination and took a motorcycle trip to Alaska, where he stayed at a monastery on a small island. Once he ran out of money, he returned to work for his previous employer, which again sent him to Afghanistan (Press 2021: 112–3). In this instance, too, Aaron's physiological discomfort was surely an indication of some form of conflict, which prompted him to withdraw and reflect. Aaron's retreat to a monastery certainly suggests that he was engaged in some form of soul-searching. But his choice to return to work as a drone pilot is an important indication that he, at this point, had not settled on an explicitly moral interpretation of his distress.

We cannot say that Aaron's interpretation of his first episode of distress, not marked by an explicit moral interpretation, is any less true than the later one. Physiological feelings do not have any moral meaning unto themselves; agents impart this meaning when they reflect on their experiences and endow them with moral significance. With time, in grappling with his distress, Aaron's search for a convincing interpretation ultimately comes to colour it through a specific lens of moral injury.

4. Emotion Perception as Decision Making

The example of Chris Aaron suggests that emotion perception may not be as straightforward and legible as most theories of emotions tend to assume. Aaron's situation is, however, rare and even extreme, and may not necessarily be representative of the way we experience our emotions on an everyday basis. I will return to address this objection in the next section, but let's bracket this concern for the moment, and consider in more detail why teleological models of emotion cannot adequately make sense of cases like Aaron's—i.e., cases where agents do not always know what they are feeling, and where their process of understanding their own emotions and physiological states may be extensive and inconclusive. My alternative reading of Aaron's case suggests that there are not distinct kinds of physiological distress with ultimate end goals, only different ways to understand our distress by making it meaningful to ourselves and others.

From this perspective, emotion perception is not about discovering what a specific feeling is really about (i.e., finding its true inherent meaning and intentionality) or what it *reveals* in the typical language of basic emotion theories (Ekman 2003). Making sense of our emotions is more akin to making an interpretive decision. As philosopher Peter Goldie puts it:

> One's own psychological states are not revealed to one in a pellucid way, especially at a time of mental turmoil: one's reasons do not come with their precise weights already established; one's feelings do not come with a precise attitude and content already in place. (Goldie 2012: 147)

The idea here is that emotion perception is not a matter of discovery (of some inherent moral trait or potential in the affective response), but rather a decision we arrive at.[14] Goldie readily concedes that the metaphor of deciding (of bestowing weights) is question-begging when we apply it to emotions, and cases of inner turmoil, like Aaron's, "where nothing is settled, least of all one's conception of oneself and one's appropriate life" (Goldie 2012: 147). In such times of conflict and confusion, it does not feel like we are able to make any kind of decisions:

---

14 Goldie borrows this idea from Richard Moran: "Sometimes, making up your mind when in a state of confusion can be not so much a matter of discovery, of finding out what you think is right, but rather a decision" (Moran 2001: 57–60, quoted in Goldie 2012: 146).

> Our agency is threatened at these times, as if we are not in control of our destiny; action seems passive, as if the surface of consciousness is being moved by some other deeper force, some tectonic plate below the surface. . . . At these times, just because one's mind is in a mess, one cannot say precisely what is going on in it. (Goldie 2012: 147)

But at some point, as Goldie puts it, "out of this mess emerges something that is, finally, more or less settled" (Goldie 2012: 148). Aaron, as we have considered, begins to understand his initial distress as the first sign of the guilt and shame he eventually comes to confront. It may certainly feel like he is discovering the true moral value of those initial feelings, but we should be cautious of echoing such autobiographical narratives and their reconstruction of events. Goldie warns that individuals construct such narratives to "explain what happened, to find agency and internal meaningfulness, precisely where it is not to be found" (Goldie 2012: 148). However productive and meaningful such narratives may be, they are constructed by us, and do not reflect some objective, tangible truth about our affective states.

The central problem that Goldie identifies is that we often end up construing our "past thoughts, feelings, and deliberations as more determinate than they in fact were" (Goldie 2012: 148). We project meaning onto our past actions, feelings, and thoughts because "we cannot bear the thought that there is no narrative explanation available of what happened in a way that provides internal meaningfulness" (Goldie 2012: 148).[15]

A consequence of this view of emotion perception is that the mere conceptualisation (articulation, understanding) of a distress response as a specific emotion is always normative. When we make sense of our emotions, the process is never purely descriptive, i.e., a matter of finding out which emotion we are feeling (Goldie 2009: 2).[16] Again, emotions are not distinct objects that are waiting to be discovered and named correctly by agents. To make sense of their emotions, agents draw on whatever hermeneutic equipment—concepts, scripts, emotion words—is available in their specific cultural, political, and social environment. The words we apply to our emotions serve as the interpretative lens that colours our experience with specific values, and motivate us accordingly.[17]

Ultimately, the fitness of an emotion concept to a situation depends on the individual agent and their broader context. As we have seen in the case of Chris Aaron, our understanding of an emotion may also change over time. The first time Aaron felt his arm go numb during the lie detector test, he withdrew from the application process but ultimately returned to Afghanistan. On this occasion, he did not end up interpreting his physiological stress with moral certainty. Or, if his moral concerns were emergent, he did not sustain them with decisive action. Later at the veteran's retreat, in a new environment, upon reflection and with a new range of interpretative frames, he settled on a moral interpretation of his stress.

---

15 See also Susan Neiman's argument that we not only expect the world to be intelligible in a theoretical sense but also require that the world is intelligible in a practical and moral sense. For this reason, we expect agents to be able to make their actions intelligible for us (Neiman 2002).

16 For a fully fleshed-out argument for why any kind of emotion perception is normative, at least in a minimal operative sense (Loidolt 2018), see Munch-Jurisic 2021; 2022: chap. 4.

17 Importantly, in many everyday emotional experiences we do not explicitly reflect on, or name, our emotional experience with words. This naming process (where we are hunting for the right words) emerges when we are asked how we feel or when we are in doubt. Nonetheless, the emotion concepts we have previously encountered and learned still serve as an implicit form of heuristics (fast forms of thinking, like biases) even though these concepts and ideas may not be phenomenologically salient in the immediate moment that we put them to use (Munch-Jurisic 2022: chap. 4). Emotion perception in human adults is therefore always mediated by the influence of the culture that surrounds them. Thank you to Laura Silva for helping me articulate this.

What this alternative reading suggests is that the ontology of what we call "emotions" cannot be separated from their epistemology. This reading is in line with the construction view of emotions, which also argues that the way in which we perceive an affective or physiological state, and which names we attribute to it, are what constitute an emotion (Barrett 2017). In the earlier discussion of reactive crying, we observed that the ability to feel vicarious distress, i.e. to respond to other people's suffering, is an emotional resonance mechanism, rooted in innate capacities. But the triggers of our affective responses—as well as who we empathise with— will depend on the values and political and social commitments that individuals internalise from their environments (Bloom 2018: 31; Munch-Jurisic 2020). Values and motivations are not, as theories like Batson's suggest, embedded in affective and physiological impulses. Instead, our innate emotional capacities serve as templates that a broad range of values, morals, and ideologies can occupy (Munch-Jurisic 2022: chaps. 4–5).

5. Disorientation as the Norm?

In this paper I propose that indeterminacy is a core part of emotion perception, and that the disorientation that Chris Aaron experiences is much more normal than standard theories of emotions assume. To be fair, traditional emotion theory certainly *can* account for disorientation and confusion in emotion perception, but such states typically only figure as exceptions, aberrations, or temporarily points of transition. This is also Goldie's perspective, whose focus is on situations of crisis where human beings have not made up their minds yet. Some readers may therefore still be sceptical. Is the physiological distress that Chris Aaron experiences not an anomaly? Do we not tend to experience emotions directly, clearly, and transparently? In fact, we often do not. In the following section, I will establish that disorientation is, if not the norm, a key component in our emotional lives.

At any given moment in time, even when we are resting and sleeping, the heart is beating, blood is pumping through our veins, and our breaths follow a certain rhythm. We are rarely consciously aware of these basic bodily functions, but they remain present, working in the background to ensure a well-functioning body. Abrupt changes in these bodily sensations will trigger our attention and prompt us to set out smaller or bigger theories about why we are feeling a particular way. A sharp sting in the stomach may be an indication of hunger, a dry mouth an indication of thirst, and so on. Researchers agree that interoception, the ability to sense and pick up on these basic bodily rhythms and cues, is fundamental for human health. Without a well-functioning interoceptive sense we risk jeopardising basic metabolic processes that are foundational for any living organism (Berntson, Gianaros, and Tsakiris 2018; Quigley et al. 2021). Interoceptive awareness is also a core feature of how we experience our own emotional life, but there are multiple ways that we can we theorise our rudimentary bodily feelings.

A couple years ago, I asked one of my students if she would be interested in presenting her independent research at a public event. For several days I did not hear back from her, and in the eventual email she declined the invitation. She had several good practical reasons for why she could not attend, but she also told me that she felt very nervous at the thought of such a performance. The feeling made her doubt her academic skills and whether she would be able to carry out such a lecture. Here we have a very ordinary example of emotional discomfort, where it is truly impossible to give an objective explanation of the student's feelings. I suggested it was completely normal to feel nervous at the prospects of a public performance and assured her that this should not be interpreted as sign that she was not fit for academia. But this suggestion was no less normative, and no more true, than her own initial interpretation. Or, put differently, the student's nervousness did not

come with an underlying meaning that could be uncovered or discovered. As Lisa Feldman Barrett would put it, we don't simply identify feelings; we construct them (Barrett 2017).

Academic life abounds with such disorienting feelings. When receiving criticism or feedback on our work, we may experience a blur of emotions such as defensiveness, self-doubt, anger, or gratitude; when we receive compliments or praise, we may also find ourselves diverging between gratitude, humility, self-doubt, or imposter syndrome. And feelings of disorientation are certainly not confined to the domain of academic work. Surely, the feeling of being nervous is one of the most ordinary and rudimentary human feelings we can think of. When we are nervous, we may begin to sweat, our stomach twinges, the heart races, and we may even begin to shake as Chris Aaron reported. But this nervousness can be understood through a wide range of interpretations. Each interpretation (or emotion concept) we apply implicitly contains a set of normative ideas about how we *should* be responding to, or acting because of, such an emotional state. In any major life transition, such as graduating from college, embarking on a new career, or getting to know a new partner, we find ourselves overwhelmed with nervousness, which prompts us to conjure tentative guesses about why we feel this way. Our attempts at grasping our emotional lives are often confused, resulting in multiple, conflicting, and contradictory interpretations, and therefore also multiple roads of actions we can embark on.

Let us return to the core focus of this article, namely the specific feeling of distress that we feel when we witness someone suffering. One of the most common such instances is the distress of new parents when they try to comfort their upset newborn, a feeling common among human beings and many other mammals. The prolonged wailing of newborns is intense and nearly impossible to ignore. Recall that the ability to feel vicarious distress is a spontaneous emotional resonance mechanism, rooted in innate capacities. In fact, one theory is that the capacity to relate to others' distress evolved exactly for the purpose of parenting (Decety and Lamm 2006; Zahavi and Rochat 2015; Preston and de Waal 2002).

What exactly are parents feeling in those tense, stressful moments when it is impossible to soothe their crying child? Is there any way we can objectively pin down what their distress really means or, in Batson's terms, what ultimate goal their emotional state motivates? Their distress is surely vicariously induced, and their intention certainly prosocial: they wish to wish to console the child. But new parents also feel tired, frustrated, even angry, inadequate as a caregiver, guilty, or ashamed before the gaze of others if they are in a public space. To what extent does it make sense to categorise this jumble of feelings—all of which are taking place simultaneously—as primarily self- or other-directed? It is not possible to obtain a bird's eye perspective on what feeling is *really* at stake in these tense moments, or to point out their one true motivational trajectory.[18] What parents eventually end up doing is the product of an active choice, often aided by advice they pick up from their surroundings. One piece of common advice, for example, is to leave the infant in the crib for a moment if one feels helpless and overwhelmed. But even if a parent construed their agitation as anger and recognised the need to step away or call for help, we should not conclude that this reveals an egoistic and antisocial affective state. We cannot assume an exclusive bond between anger and certain motivational tendencies (Cameron, Lindquist, and Gray 2015; Silva 2021a).

In the various cases discussed above, though our conflicting and disorienting feelings may pull us in different directions, we are never at their mercy. Our emotions may be puzzling and difficult to decipher, even conflicting, but they are not distinct entities waiting to be discovered. The interpretations we arrive at, our

---

18 As plenty of research has made clear, vicarious distress in and of itself does not merely motivate prosocial behaviour (Munch-Jurisic 2022: chap. 2).

expressions and explanations to friends, are a sort of choice, an emphasis on one set of explanations in favour of another. Of course, we need shared concepts, names, and words to communicate with and understand each other—"Good fences make good neighbors," the neighbour in Frost's poem maintains—but the concepts that have been so defined (e.g. happy, sad, angry, disgusted) do not refer to distinct ontological or natural categories (Barrett 2017). They are shared, hermeneutic resources with which we cultivate our emotional well-being and a well-functioning social world.

6. The Value of Disorientation

Something there is that does not love a wall, which defeats our fences and resists our attempts at categorisation. In many everyday situations, we are left wondering what we feel and why. This lack of determinateness is a fundamental, constitutive feature of our emotional lives. It can inspire us to think of new and better ways to conceive of ourselves and reconfigure the way we see the world around us (Silva 2021b; 2022; Harbin 2016; Jaggar 1989). The concept of moral injury, a source of relief for Chris Aaron and many other veterans, is but one such example.

The primary aim of this paper has been to criticise the ontology presented in theories of emotions that rely on a teleological framework. But this critique of traditional emotion theory extends beyond questions of ontology and categorisations. It is also a disagreement about how we should theorise human nature. Traditional emotion theory is built on moments where we know our feelings—we have found the right categories, words, or heuristics to express what we feel. But human emotions, as we all know, are messy. To ignore this fact, and to fail to account for emotional disorientation as a significant part of our emotional life, is theoretically untenable.

Even cases like Aaron's should not be considered a deviation from the norm. Many veterans find themselves in a similar situation as Aaron. As anthropologist Tine Molendijk found in semi-structured interviews of Dutch veterans who also experienced moral injury:

> Some veterans explicitly expressed uncertainty or confusion about the significance of their experiences. These veterans said that they "cannot work it out" and that they "cannot resolve it." Others expressed uncertainty or confusion more implicitly (perhaps unconsciously), by uttering ambivalent or even conflicting interpretations of their experiences. (Molendijk 2018: 5–6)

Molendijk emphasizes that the veterans' confusion should not be "readily (or only) be conceived of as a disorder, but rather (or also) as an as an ethical struggle" as a result of crushed moral certainties (Molendijk 2018: 7).[19]

The disorientation of veterans is not just normal and widespread,[20] it reflects well on them as moral agents (Kurth 2018a). In cases where veterans either perpetrated or witnessed others commit actions that violated their moral integrity, it would be chilling if they *were* able to uphold a stable self and not enter a state of disorientation. Sometimes we should not be able to sleep at night, and it is ethically prudent to be riddled with anxiety and uncertainty (Kurth 2018b: chaps. 4–6). A full discussion of the moral value of disorientation must

---

19 There is no doubt that Aaron and many other veterans in similar situations often need intense medical and therapeutic help, but this does not exclude the possibility that their struggles also entail an ethical component.

20 The literature covering moral injury among military veterans is vast and in rapid development. See Litz 2023: 20 for a recent literature review.

be left to another occasion. But the mere capacity to enter a state of disorientation—to be in doubt and second-guess our own assumptions—must be considered a core feature of human nature, not an anomaly.[21]

To be sincere, as Goldie writes, we must take disorientation seriously and refuse the determinate alternative (Goldie 2012: 147).[22] Goldie's point here is twofold. On an individual level, this requires us to embrace and take responsibility for the eventual "decisions" we make about our affective lives. But, from a theoretical perspective, the point is that there is no objective narrative or explanation that corresponds completely to our emotions. When teleological theories posit that emotions be captured and defined in such determinate terms, they give a mistaken account of our emotional life.

The problem is not just that teleological theories argue for determinate labels and distinct emotion categories, but also the implicit suggestion that we can predict how human beings will act based on which (distinct) affective states they are in. The core problem in teleological theories of emotions is the implicit assumption that it is possible to explain the behaviour of human beings in finite, definite terms. But, as Hannah Arendt pointed in *The Human Condition*, it is simply not possible to define human nature as we define other entities surrounding us. To attempt such an endeavour is like trying to jump over our own shadow (Arendt 1998: 10).

From Arendt's perspective, human beings can never know themselves fully. She developed this anti-essentialist and anti-teleological thesis of human nature as a criticism of the behaviourism of her time. She draws on Saint Augustine, who writes his confessions to try to understand his sinful youth before conversion to Christianity. "A question have I become to myself," he ponders, but Augustine does not regard this lack of complete self-comprehension as a problem (Arendt 1998: 10). It is rather something to marvel at.[23]

For Arendt, too, the mystery of self-knowledge is something equally terrifying and miraculous. Her entire political philosophy builds on the idea that human beings are only able to act freely because of this "startling unexpectedness," which is "inherent in all new beginnings and in all origins" (Arendt 1998: 178). Freedom is not just choosing between a certain number of alternatives; freedom is the ability to present a new perspective that is unexpected and unpredictable. Without this capacity, we would not be able to challenge the prevailing status quo. And veterans like Aaron would not be able to come up with new perspectives on why they feel distressed.

To a significant extent, we might say that agents themselves are not fully conscious of "who" they are. In fact, Arendt argues, "it is more likely that the 'who', which appears so clearly and unmistakably to others, remains hidden from the person himself" (Arendt 1998: 179). This lack of transparency within a context of self-interpretation becomes especially evident when we consider cases of physiological stress such as that experienced by Aaron.[24] It was not until Aaron entered a community of fellow veterans with similar experiences that he was able to articulate and arrive at a clearer understanding of what he felt. From Arendt's perspective, this is because agents only come to know who they are through their interactions with others.[25] Their understanding of themselves is moulded by their surroundings and the hermeneutic equipment that is

---

21 For discussions on the moral and political value of discomfort, see for example Harbin 2016; Kurth 2018b.

22 Here Goldie is quoting Stuart Hampshire (1972: 243).

23 Augustine writes: "Nor do I myself comprehend all that I am. Therefore is the mind too strait to contain itself. And where should that be, which it containeth not of itself? Is it without it, and not within? How then doth it not comprehend itself? A wonderful admiration surprises me, amazement seizes me upon this" (St. Augustine n.d.: bk. X chap. VIII).

24 See also Goldie's discussion of this (2012: 136) in which he draws on Crispin Wright (1998: 16).

25 For a recent experimental exploration of a similar thesis, see Bolis and Schilbach 2020.

made available for their use.[26] When we participate in conversation with others, we are asked to give reasons for our actions, and to stand accountable for our conduct (Arendt 1998: 182). It is in such moments, where we explain ourselves, that we can begin to name our feelings and try to reconstruct what they meant. But the way we construe our feelings is no finite process, through which we eventually discover the correct, determinate label, to assign to them. Our emotional life does not have a pre-defined telos. How we end up construing our emotions may even surprise ourselves. It is this indeterminate feature of human nature that makes moral transformation possible. Disorientation is not just a normal feature of human emotional life; it is also a constituent part of moral agency.

7. References

Antić, A. 2016. *Therapeutic Fascism: Experiencing the Violence of the Nazi New Order*. 1st ed. Oxford University Press.

Arendt, H. 1998. *The Human Condition*. 2nd ed. University of Chicago Press.

Barrett, L. F. 2017. *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.

———. 2022. "Context Reconsidered: Complex Signal Ensembles, Relational Meaning, and Population Thinking in Psychological Science." *The American Psychologist* 77 (8): 894–920.

Batson, C. D. 2009. "These Things Called Empathy." In *Social Neuroscience of Empathy*, edited by J. Decety and W. Ickes, 3–15. MIT Press.

———. 2014. "Empathy-Induced Altruism and Morality: No Necessary Connection." In *Empathy and Morality*, edited by H. L. Maibom, 41–58. Oxford University Press.

Batson, C. D., J. Fultz, and P. A. Schoenrade. 1987. "Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences." *Journal of Personality* 55 (1): 19.

Berntson, G. G., P. J. Gianaros, and M. Tsakiris. 2018. "Interoception and the Autonomic Nervous System: Bottom-Up Meets Top-Down." In *The Interoceptive Mind: From Homeostasis to Awareness*, edited by M. Tsakiris and H. De Preester, 3–24. Oxford University Press.

Betz, N., K. Hoemann, and L. F. Barrett. 2019. "Words Are a Context for Mental Inference." *Emotion* 19 (8): 1463–77.

Bloom, P. 2018. *Against Empathy: The Case for Rational Compassion*. Vintage.

Bolis, D. and L. Schilbach. 2020. "'I Interact Therefore I Am': The Self as a Historical Product of Dialectical Attunement." *Topoi* 39 (3): 521–34.

Boudreau, T. E. 2011. "The Morally Injured." *The Massachusetts Review* 52 (3/4): 746–54.

Cameron, C. D., K. A. Lindquist, and K. Gray. 2015. "A Constructionist Review of Morality and Emotions: No Evidence for Specific Links Between Moral Content and Discrete Emotions." *Personality and Social Psychology Review* 19 (4): 371–94.

Decety, J. and C. Lamm. 2006. "Human Empathy through the Lens of Social Neuroscience." *The Scientific World Journal* 6: 1146–63.

Ekman, P. 1992. "An Argument for Basic Emotions." *Cognition & Emotion* 6 (3–4): 169–200.

———. 2003. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Times Books.

Frost, R. 1914. *North of Boston*. David Nutt

Goldie, P. 2009. *The Emotions: A Philosophical Exploration*. Clarendon Press.

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

Hampshire, S. 1972. *Freedom of Mind and Other Essays*. Clarendon Press.

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

---

26 Arendt calls this *plurality*. It is the already existing web of relationships and systems for organisation and interpretation that we are thrown into (Arendt 1998: 184). For contemporary theoretical approaches that advocate the influence of the environment on our epistemic resources, see Haslanger 2020; McGeer 2007a; 2007b; Zawidzki 2013; Munch-Jurisic 2022: chap. 4.

Haslanger, S. 2020. "Cognition as a Social Skill." *Australasian Philosophical Review* 3 (1): 5–25.

Hoemann, K., Z. Khan, M. Feldman, C. Nielson, M. Devlin, J. Dy, L. F. Barrett, J. B. Wormwood, and K. Quigley. 2020. "Context-Aware Experience Sampling Reveals the Scale of Variation in Affective Experience." *Scientific Reports* 10: 12459.

Hoffman, M. L. 2000. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge University Press.

Hutto, D. D., I. Robertson, and M. D. Kirchhoff. 2018. "A New, Better BET: Rescuing and Revising Basic Emotion Theory." *Frontiers in Psychology* 9.

Jaggar, A. M. 1989. "Love and Knowledge: Emotion in Feminist Epistemology." *Inquiry* 32 (2): 151–76.

Keltner, D., J. L. Tracy, D. Sauter, and A. Cowen. 2019. "What Basic Emotion Theory Really Says for the Twenty-First Century Study of Emotion." *Journal of Nonverbal Behavior* 43 (2): 195–201.

Kurth, C. 2018a. "Anxiety: A Case Study on the Value of Negative Emotion." In *Shadows of the Soul*, edited by C. Tappolet, F. Teroni, and A. Konzelmann Ziv. Taylor & Francis.

———. 2018b. *The Anxious Mind: An Investigation into the Varieties and Virtues of Anxiety*. MIT Press, 95-104.

———. 2022. *Emotion*. Routledge.

Lazarus, R. S. 1991. *Emotion and Adaptation*. Oxford University Press.

Lench, H. C., S. A. Flores, and S. W. Bench. 2011. "Discrete Emotions Predict Changes in Cognition, Judgment, Experience, Behavior, and Physiology: A Meta-Analysis of Experimental Emotion Elicitations." *Psychological Bulletin* 137 (5): 834–55.

Leys, R. 2017. *The Ascent of Affect: Genealogy and Critique*. University of Chicago Press.

Litz, B. T. 2023. "The Future of Moral Injury and Its Treatment." *Journal of Military, Veteran and Family Health* 9 (2): 1–5.

Litz, B. T. and P. K. Kerig. 2019. "Introduction to the Special Issue on Moral Injury: Conceptual Challenges, Methodological Issues, and Clinical Applications." *Journal of Traumatic Stress* 32 (3): 341–9.

Litz, B. T., N. Stein, E. Delaney, L. Lebowitz, W. P. Nash, C. Silva, and S. Maguen. 2009. "Moral Injury and Moral Repair in War Veterans: A Preliminary Model and Intervention Strategy." *Clinical Psychology Review* 29 (8): 695–706.

Loidolt, S. 2018. "Experience and Normativity: The Phenomenological Approach." In *Phenomenology and Experience*, edited by A. Cimino and C. Leijenhorst, 150–65. Brill.

Maibom, H. L. 2014a. *Empathy and Morality*. Oxford University Press.

———. 2014b. "Introduction: (Almost) Everything You Ever Wanted to Know about Empathy." In *Empathy and Morality*, edited by H. L. Maibom, 1–40. Oxford University Press.

McGeer, V. 2007a. "The Moral Development of First-Person Authority." *European Journal of Philosophy* 16 (1): 81–108.

———. 2007b. "The Regulative Dimension of Folk Psychology." In *Folk Psychology Re-Assessed*, edited by D. D. Hutto and M. Ratcliffe, 137–56. Springer Netherlands.

Molendijk, T. 2018. "Toward an Interdisciplinary Conceptualization of Moral Injury: From Unequivocal Guilt and Anger to Moral Conflict and Disorientation." *New Ideas in Psychology* 51 (December): 1–8.

Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.

Munch-Jurisic, D. M. 2020. "Against Comfort: Political Implications of Evading Discomfort." *Global Discourse: An Interdisciplinary Journal of Current Affairs* 10 (2): 277–97.

———. 2021. "Lost for Words: Anxiety, Well-Being, and the Costs of Conceptual Deprivation." *Synthese* 199 (5): 13583–600.

———. 2022. *Perpetrator Disgust: The Moral Limits of Gut Feelings*. Oxford University Press.

Neiman, S. 2002. *Evil in Modern Thought: An Alternative History of Philosophy*. Princeton University Press.

Press, E. 2021. *Dirty Work: Essential Jobs and the Hidden Toll of Inequality in America*. Farrar, Straus and Giroux.

Preston, S. D. and F. B. M. de Waal. 2002. "Empathy: Its Ultimate and Proximate Bases." *The Behavioral and Brain Sciences* 25 (1): 1–20.

Quigley, K. S., S. Kanoski, W. M. Grill, L. F. Barrett, and M. Tsakiris. 2021. "Functions of Interoception: From Energy Regulation to Experience of the Self." *Trends in Neurosciences* 44 (1): 29–38.

Scarantino, A. 2014. "The Motivational Theory of Emotions." In *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*, edited by J. D'Arms and D. Jacobson, 156–85. Oxford University Press.

Scarantino, A. and P. Griffiths. 2011. "Don't Give Up on Basic Emotions." *Emotion Review* 3 (4): 444–54.

Silva, L. 2021a. "The Efficacy of Anger: Recognition and Retribution." In *The Politics of Emotional Shockwaves*, edited by A. Falcato and S. G. da Silva, 27–55. Springer International Publishing.

———. 2021b. "The Epistemic Role of Outlaw Emotions." *Ergo* 8.

———. 2022. "On Being Angry at Oneself." *Ratio* 35 (3): 236–44.

Simner, M. L. 1971. "Newborn's Response to the Cry of Another Infant." *Developmental Psychology* 5 (1): 136–50.

St. Augustine. n.d. "Confessions of Saint Augustine." *Christian Classics Ethereal Library*. Accessed May 16, 2023. https://www.ccel.org/ccel/augustine/confess.xi.viii.html.

Wiinikka-Lydon, J. 2017. "Moral Injury as Inherent Political Critique: The Prophetic Possibilities of a New Term." *Political Theology* 18 (3): 219–32.

———. 2019. "Mapping Moral Injury: Comparing Discourses of Moral Harm." *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 44 (2): 175–91.

———. 2022. "Critiquing the Subject of Moral Injury." *Journal of Military Ethics* 21 (1): 39–55.

Wright, C. 1998. "Self-Knowledge: The Wittgensteinian Legacy." In *Knowing Our Own Minds*, edited by C. Wright, B. C. Smith, and C. Macdonald, 101–22. Oxford University Press.

Zahavi, D. and P. Rochat. 2015. "Empathy ≠ Sharing: Perspectives from Phenomenology and Developmental Psychology." *Consciousness and Cognition* 36 (November): 543–53.

Zawidzki, T. W. 2013. *Mindshaping. A New Framework for Understanding Human Social Cognition*. MIT Press.