Research Articles

EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

Open access

TABLE OF CONTENTS

RESEARCH ARTICLES

ARE THERE "MORAL" JUDGMENTS?

David Sackris[1] and Rasmus Rosenberg Larsen[2]

[1] Arapahoe Community College
[2] University of Toronto, Mississauga

ABSTRACT

Recent contributions in moral philosophy have raised questions concerning the prevalent assumption that moral judgments are typologically discrete, and thereby distinct from ordinary and/or other types of judgments. This paper adds to this discourse, surveying how attempts at defining what makes moral judgments distinct have serious shortcomings, and it is argued that any typological definition is likely to fail due to certain questionable assumptions about the nature of judgment itself. The paper concludes by raising questions for future investigations into the nature of moral judgment.

Keywords: *metaethics; moral judgment; judgment; ontology.*

Introduction

An enduring assumption in the Western philosophical tradition is that moral judgment—namely, judgments about *rightness* and *wrongness*—constitutes a specific type of judgment, distinct from other judgment types. In other words, philosophers have traditionally assumed that when human beings make 'moral' judgments, they are doing something typologically (i.e., categorically) different from when they make judgments about other affairs, such as judging that today is Tuesday, that the moon is full, or that 2+3=5.

As noted by Walter Sinnott-Armstrong and Thalia Wheatley (2012; 2014), this assumption—that moral judgments form a distinct type of judgment–has been at the center of various conversations in the history of moral philosophy. For example, Immanuel Kant is often interpreted as proposing in the *Groundwork of the Metaphysics of Morals* (1998) that a moral judgment is made by a distinct faculty of the mind, and that moral judgments have distinctive properties that other types of judgments lack: only judgments of morality are categorically binding (see also Stich 2006).[1]

Most contemporary philosophers seemingly continue to maintain this assumption, as exemplified by the questions that dominate current debates, such as: are moral judgments made using cognitive or non-cognitive processes? Are moral judgments necessarily motivating? Can moral judgments be true? If moral judgment was not assumed to be a distinctive judgment type, then it is not obvious that these questions would at all be meaningful (for an overview of the sorts of questions focused on in contemporary metaethics, see Sayre-McCord 2014; Smith 1994). Note that we do not usually ask such questions about many other judgments. That is, we do not ask if judgments about the weather necessarily *motivate*, or if such a judgment is true in the *same way* as other judgments. We don't seem to presuppose that there is something typologically distinctive about many of the judgments we make in our day to day lives. They are just, well, judgments.

While it may be intuitively appealing to assume that moral judgments are typologically distinct, the assumption arguably presents challenges that have been scarcely recognized by contemporary philosophers. One such challenge is rather fundamental, as it has to do with providing a clear and

---

[1] See Sackris and Larsen (2022) for an overview of contemporary metaethics scholarship and the widely held commitment to the position that moral judgment constitutes a distinctive type. See also the anthology by Decety and Wheatley (2015) for contemporary approaches, and Verplaetse (2009) for a historical overview of the (neuro) science of moral judgment.

unambiguous definition of what makes a judgment 'moral'. That is, if moral judgment constitutes a distinct type of judgment, then we should be able to state what features distinguish this type of judgment from other types of judgments. [2] Sinnott-Armstrong and Wheatley, for example, endorse such a definitional approach when they state, "A group of things are unified in the relevant way if and only if they share some feature that enables important universal generalizations about its distinctive properties" (2012, 356). This sort of definitional approach can be written out as the logical formula, *A is a B that Cs*, meaning that we must be able to finish the sentence: *Moral judgments are judgments that 'C',* where '*C*' is filled in by some significant feature that all moral judgments share but that is not possessed by other, non-moral judgments (also known as an *Aristotelian* definition, and this form constitutes the basic logic behind scientific taxonomies; see Arp, Smith, and Spear 2015; Berg 1982; Seppälä, Ruttenberg, and Smith 2017).

In this paper we aim to build on prior arguments that have sought to cast doubt on the idea that moral judgments form or instantiate a unified type, whereby all instances of moral judgments have a shared set of features (i.e., Sinnott-Armstrong and Wheatly 2012; 2014; Stich 2006; Sackris and Larsen 2021). Among these contributions, some have questioned the unity of the category of moral judgement (e.g., Sinnott-Armstrong and Wheatley 2012), but so far none have explicitly argued that the category itself must be rejected. Here we provide and test such an argument, namely: that it is not only difficult, but probably also impossible, to define what makes a judgment of the 'moral' type (i.e., defining the '*C*'). We contend that common definitions of moral judgment do not sufficiently distinguish a 'moral' judgment from other judgments and/or judgment simpliciter. We advance this proposition by first pointing out that the term 'judgement' is itself lacking a proper definition, whereafter we critically survey three possible accounts of 'moral' judgment: (1) we consider the common position that moral judgment is distinguished from other judgment types by its content; (2) we consider the possibility that moral judgments are distinguished from other judgments by brain processes; and lastly (3) we consider one of the few explicit and contemporary defenses of the position that moral judgments constitute a distinctive kind (i.e., Kumar 2015; 2016a; 2016b). We show that all three proposals fail to meaningfully differentiate 'moral' judgments from other judgment types. Finally, we

---

[2] Here we use 'type' and 'kind' interchangeably. We do not have any philosophical commitments as to what constitutes a 'natural kind' (for review, see Brzović 2018), nor do we think such nuances will undermine the main argument articulated in this paper. We use 'kind' and 'type' to mean a distinctive category that is distinguished by some shared property or feature (or set of shared properties or features).

conclude by raising questions for future investigations into the nature of moral judgment.

Before we proceed with our argument, a point of clarification is necessary. We do not mean to argue that the phrase 'moral judgment' has no *meaning* (i.e., that the phrase is semantically empty). To most people, it is undoubtedly meaningful when a person utters "I find smoking immoral" or "It is immoral to be so selfish". When terms like 'moral' and 'immoral' are used as adjectives, these can convey all sorts of meanings, for example, that one approves or disapproves of certain behaviors. The argument we promote in this paper aims to raise doubt about the position on which such judgments have a distinct typological *referent*, or whether they refer to something that constitutes a distinctive type (i.e., that it is meaningfully distinguishable from what we might intuitively think of as other judgment types). Indeed, what we aim to reject is the idea that all judgments typically referred to as 'moral' necessarily have shared features that allow us to infer that they instantiate a universal sub-category distinct from other judgment types. On our view, when people call something a 'moral judgment', there is no clear corresponding referent (such as a specific cognitive type or process). The argument we provide here is an attempt to show that there are good reasons to reject the idea of 'moral judgment' as a distinctive type.

1.   What is a Judgement?

What we see as a key hindrance in the project of defining 'moral' judgment is that it is unclear what exactly is meant by the term 'judgment'. Ostensibly, any attempt to define 'moral judgment' should begin with defining what 'judgment' is, and only thereafter define what differentiates a 'moral' judgment from ordinary and/or other types of judgment (i.e., as captured in *A is a B that Cs*) (e.g., Arp, Smith, and Spear 2015; Berg 1982; Seppälä et al. 2017). For example, when biologists define and taxonomize newly discovered novel-looking organisms, they must first determine whether it is, say, a vertebrate animal, and only thereafter whether it is a mammal, reptile, bird, etc. If it is a mammal, then it must be determined whether the animal falls under any of the various species-categories in the mammalian taxonomy or whether it can be genuinely distinguished from its evolutionary siblings and ancestors, namely, the 'C' in *A is a B that Cs*. Consider, for example, recent controversy concerning whether 'Brontosaurus' and 'Apatosaurus' constitute two distinctive species of dinosaur. Undoubtedly, extent fossils indicate that the animals in question are part of the diplodocid family, but we only have two distinctive species if we can find some feature that reliably distinguishes one set of bones from another (Osterloff, n.d.).

However, in contemporary moral philosophy, this standard procedure in taxonomic definitions has never really been strictly applied. For example, in Richmond Campbell's 2007 paper entitled, "What is Moral Judgment?", there is no discussion of what a 'judgment' simpliciter might be (i.e., there is no typological definition of judgment: the 'B'). Whether all judgments have anything in common, or if there are several different types of judgment in general is not discussed. Campbell does state that "Moral judgments are (or express) states of belief" (2007, 321), but says nothing about what makes a 'moral' judgment different from other types of judgments. Indeed, speaking to Campbell's claim concerning 'moral' judgments, we might ask whether or not it is true that *all* judgments express states of belief?[3] If so, then his definition fails to differentiate anything distinctive.

Similarly, in a 2015 paper by Victor Kumar entitled, "Moral Judgment is a Natural Kind", the first line of the paper reads: "Moral judgments seem to be different from other normative judgments, even apart from their characteristic subject matter" (2015, 2887). In this paper, Kumar is distinguishing moral judgment in part by its supposedly distinctive subject matter, but like Campbell, there is no clear attempt at defining 'judgment', let alone what a 'normative' judgment is (see also May and Kumar 2018).

Further examples come to mind. In Joshua Glasgow's (2013) exploration of the phenomenological essence of moral judgment, there is no discussion of judgment itself. Even in Michael Smith's modern classic, *The Moral Problem*, in which he characterizes "the central organizing problem in contemporary metaethics" (1994, 11) as an inability to satisfactory define moral judgment (given the commitment to a Humean account of motivation), there is little discussion of what 'judgment' might be. Smith does hold that moral judgments necessarily motivate, but he does not clarify whether moral judgments are the *only* judgments that necessarily motivate (see especially chapter 3).

In these contemporary publications, it would appear as if the reader is supposed to have a kind of fundamental or intuitive understanding of what a 'judgment' is; and thereby also, what a 'moral' judgment is. But do they? Apparently, there are good reasons to be skeptical about philosophers intuitively knowing or agreeing on what a 'judgment' is, or even supposing that the term 'judgment' picks out a single, uniform cognitive process. In

---

[3] Campbell goes on to add that the view that moral judgments are beliefs is in conflict with other beliefs that we might have about moral judgments, as well as beliefs we might have about belief itself, so even if the statement "Moral judgments are (or express) states of belief" was a kind of definition, it wouldn't be a widely accepted one. Similarly, Smith's (1994) is focused on this same problem.

Wayne Martin's historical survey of philosophical theories of judgment he concludes:

> In large part, the history I recount is a history of philosophical failure. In each study I show how seemingly promising approaches to judgment led more or less directly to theoretical impasse. The problem of judgment, it turns out, proves remarkably resistant to solution—even across a diverse range of disciplines and methodologies. (Martin 2006, 7)

We similarly propose that it would be too hasty to assume that philosophers know what makes a judgment 'moral'. *Prima facie*, what we call 'judgments' likely result from a diverse set of cognitive processes, and those processes may or may not result in end products that have shared features. Therefore, there is legitimate and fundamental reason for suspicion when the task of defining 'judgment' is itself ignored in the literature that aims at defining 'moral judgment'.

What is arguably missing in the field is a clear definition along the following lines: A 'moral' judgment is a 'judgment' that 'Cs'. And it may be speculated that this lack of a clear definition of 'moral judgment' is rooted in the lack of a clear understanding and shared consensus about how to define judgment itself. If judgment is not well understood or if we have unfounded assumptions concerning human judgment, it is going to be that much harder to successfully differentiate 'moral' judgments from other judgment types.

It is to such attempts at defining 'moral' judgment—the 'Cs'—that we now turn.

2.   Defining Moral Judgment on the Basis of its Content

In this section, we consider what we shall refer to as the *content approach*, namely, those attempts that aim to define moral judgment by stating what moral judgments are *about*.[4] That is, when a human being has made a judgment, what makes that judgment 'moral' has to do with the content of said judgment. On this approach, the definitional schema (A is a B that Cs) would be filled out in the following way: Moral judgments are judgments about a certain content.

---

[4] The content approach is also found in aesthetics. See for example Carroll (2012; 2015); cf. Sackris and Larsen (2019; 2023).

Several philosophers seem to make just this claim. For instance, Kumar partially endorses such a criterion by referencing a "characteristic subject matter" of moral judgments (2015, 2887). Another contemporary philosopher who seems to venture in this direction is Jesse Prinz, when he defines moral judgments as judgments about those *behaviors* we take to be moral (2007, 47-49). Smith appears to be saying something similar, as he believes platitudes from everyday discourse surrounding morality and moral judgment can be used to define it:

> To say that we can analyze moral concepts, like the concept of being right, is to say that we can specify which property the property of being right is by reference to platitudes about rightness. (Smith 1994, 39)[5]

The idea seems to be that we have a commonsense understanding as to what moral judgments are typically *about*.

Although the content approach for differentiating 'moral' judgment from other types of 'judgment' is certainly appealing, we believe such a definition is both problematic and fundamentally implausible.

First, many philosophers have readily admitted that it is difficult to define what exactly moral philosophy is about (i.e., its content). G.E. Moore famously claimed that 'good' itself is undefinable (1903). More recently, James Dreier states "We should just admit that it may be vague whether a given judgment is moral or not" (1996, 411). Owen Flanagan highlights the difficulty of defining (*a priori*) what features of actions make them moral issues (1993, 17). Similarly, Shafer-Landau (2015) also suggests that 'morality' is undefinable. [6] These remarks suggest that moral judgments do *not* have a "characteristic subject matter" if there is well-recognized vagueness as to whether a given judgment counts as a moral one.

---

[5] Smith does say "there are limits on the kind of content a set of requirements can have if they are to be moral requirements at all, as opposed to requirements of some other kind" (1994, 40). Of course, Smith doesn't say what those limits are exactly, and that is just our point. If a person can consider almost any issue a moral one, it is hard to see how far commonplaces about morality will get us to a meaningful distinction between the 'moral' and the 'non-moral'.

[6] See his "Introduction". Additional examples can be given. Richardson (2018), in his *Stanford Encyclopedia of Philosophy* entry on moral reasoning states "[W]e will need to have a capacious understanding of what counts as a moral question. For instance, since a prominent position about moral reasoning is that the relevant considerations are not codifiable, we would beg a central question if we here defined 'morality' as involving codifiable principles or rules". Svavarsdottir admits that "it is of course notoriously difficult to say what distinguishes moral judgments from other evaluative or normative judgments" (1999, footnote 6).

Perhaps it is difficult to define morality based on what it is *about* precisely because it appears that almost any behavior can be moralized if persons or a given society decides to moralize it (e.g., Fiske and Rai 2014; Haidt 2012): the problem here is that if any content can be identified as the subject of a moral judgment, then it would be futile to search for content that can serve as a unique identifier. As Sinnott-Armstrong and Wheatly point out:

> The mere fact that people tend to group a set together under a name does not show that there is any single feature that enables significant generalizations about all and only things in that group. (Sinnott-Armstrong and Wheatly 2014, 457)

Given the great diversity of content (e.g., behaviors, events, etc.) that have been moralized by human beings (e.g., dietary restrictions, honor killings, sexual relations; see, Fiske and Rai 2014) it is hard to see how studying this hodgepodge could tell us in *what way* moral judgments differ significantly from the judgment of other behaviors and events. If any human activity or event can serve as the basis-content for a moral judgment, then it follows that moral judgments cannot be distinguished from other judgments on the content approach.

Recognizing this line of criticism, a proponent of the content approach might try to refine their view and settle on something like 'On the surface moral judgments may seem to be about a disunified class of human activities, but actually they are unified by a focus on preventing harm'.[7] Claims like this, however, force us to look for harms to accompany *every* moral judgment, and it often doesn't matter if the supposed harms are dubious. Judgments about the wrongness of masturbation, cannibalism, sex outside of wedlock, or homosexual relations do not appear to have anything to do with harm. Finding who is supposed to be harmed by such behaviors requires a good deal of mental gymnastics. Yet, they are all examples of behaviors (i.e., content) that has or currently is being moralized.

The defender of the content approach may also try to fall back on vagueness. For example, there are many vague terms in use (e.g., baldness, heap), which philosophers typically acknowledge as less problematic than they may appear; and perhaps 'moral judgment' is just such a vague term. Just as we know a bald person when we see one, we easily recognize core

---

[7] See Sinnott-Armstrong and Wheatly (2014, 457-459) for a more thorough discussion of this point, and an able refutation.

cases of moral judgment. That there are borderline or questionable cases doesn't show that we lack a good definition of what is 'moral' content.

To fall back on vagueness seems appealing, but it is unsatisfying for at least one central reason: there is scarce agreement on what counts as 'core' cases of moral content, as persons and cultures consider different and non-overlapping contents as 'core' instances of moral judgment. Even with stereotypical moral cases such as 'murder is immoral' we find plenty of examples where cultures and persons believe that some acts that others would consider 'murder' can be morally justified. For example, genocides have been seen by their perpetrators as morally virtuous acts (e.g., Fiske and Rai 2014). When we use a term such as 'bald' to refer to a person without any hair, no one would disagree that we are using the term appropriately. It is the boundary cases that people disagree about, such as when a person is in the process of losing their hair. Falling back on vagueness is therefore not a convincing defense for the content approach since there is disagreement concerning the 'core' cases.

More significantly, we may legitimately wonder why someone would believe that it is possible to define a judgment type by its content in the first place. To attempt to define a judgment type by what is being judged seems to be the wrong way of approaching things. If 'moral' judgment really is a distinctive type of judgment, it would seem that we would want to identify the distinctive brain or mental process *first*, and then look for what kinds of things trigger *that* or *those* processes. As an analogy, we might be able to catalog items that typically cause people to become sexually aroused, but we would not want to identify or define sexual arousal with those items—since, after all, sexual arousal is a complex neurological and mental process—and we would likely admit that a person could, in theory, become sexually aroused by anything. There may be reliable triggers in large parts of any given population, but sexual arousal is not defined by these triggers.

3.   Defining Moral Judgment on The Basis of Its Brain Processes

In this section, we consider whether moral judgments can be distinguished on the basis of typologically distinct cognitive processes. That is, if a judgment is understood as the result of a cognitive *process*, it then seems possible that this process itself could be sufficiently different from other processes when making a *moral* judgment. One way to pursue this strategy, it seems, is to find seemingly uncontroversial issues of *rightness* and *wrongness* (e.g., charity, murder, etc.) and investigate how people judge *those* issues, and whether there are similar or different cognitive processes

undergirding these judgments. If there is, this would perhaps be a good reason to think that moral judgment is a distinctive kind of judgment. If there is not, this would be a reason to doubt that moral judgment is a distinctive judgment type.

In studies conducted in 2001 and 2004, Joshua Greene and others used neuroimaging technology to scan the brains of individuals considering standard moral dilemmas such as the *Trolley problem* and *Footbridge problem* (from Foot 1967; Thomson 1976). Although Greene's stated goal was to debunk the belief that deontological moral judgments were the result of pure reasoning processes (2008, 36), his findings indicate that moral judgments are not reached via a single area/system of the brain or a unified set of brain processes. In Greene's study, when subjects considered dilemmas that directly involved themselves, what Greene refers to as 'personal' moral dilemmas (e.g., would you shove one person from a bridge to stop a trolley from killing five people), areas of the brain that are believed to be involved with emotion-processing were activated. When subjects considered more 'impersonal' moral dilemmas (e.g., flipping a switch from a distance to divert a trolley from killing five and as a result killing one person instead), their judgments appeared to be correlated with activity in areas of the brain responsible for (non-emotional) cognitive processes:

> Contemplation of personal dilemmas produced relatively greater activity in three emotion-related areas: the posterior cingulate cortex, the medial prefrontal cortex, and the amygdala […]. At the same time, contemplation of impersonal moral dilemmas produced relatively greater neural activity in two classically "cognitive" areas of the brain, the dorsolateral prefrontal cortex and inferior parietal lobe. (Greene 2008, 43-44)

Greene did not initially take his findings to show that moral judgment cannot be meaningfully classified as a distinct category of judgment; it is only later that he has come to embrace such a conclusion, stating: "I believe that moral judgment is not a natural kind at the cognitive level" (2015b, 40).

Although it is excruciatingly difficult to identify functional areas of the brain with high confidence (e.g., Marek et al. 2022), what is important for our purposes is that Greene et al. (2001; 2004) showed that different areas of the brain were functioning when confronting different kinds of moral scenarios and/or problems. This reliance on different areas of the brain for forming judgments about what many would readily classify as 'moral'

dilemmas should give us pause. If different functional areas of the brain are being relied on to make judgments in these cases, then this should at least be reason enough to begin questioning our confidence in the assumption that there is a *single* brain process or system that can be identified with the term 'moral judgment'. In fact, the opposite appears to be true if we follow Greene's findings: different areas of the brain will be called on to form a judgment depending on the context. If there is no single process (or pattern of processes) that can be identified when test subjects make judgments that are assumed to be moral, then brain processes cannot serve as the basis for typologically differentiating moral judgments from other judgment types. Furthermore, if different brain processes are active during the 'moral' judgment process, this might give us reason to think that the resulting judgments from these different areas of the brain might well have different properties (e.g., beliefs in some cases, desires in others).

Building on Greene, Fiery Cushman and Liane Young (2011) have found that patterns of moral judgment can be attributed in part to regions of the brain responsible for the attribution of intentions and causation, general reasoning process that might be engaged in a variety of judgment types. From this they conclude that our moral judgments are "derived" from more general judgment forming processes (2011, 1053). Their findings suggest that there is no distinct 'moral' judgment type, but that instead the brain employs a general reasoning process that it applies to a diversity of issues. When reviewing the evidence considered in both Greene (2008) and Cushman and Young (2011), Borg and colleagues (2011) reached a similar conclusion:

> Consistent with their structure in the deep brain and given their participation in negative judgments in many contexts, the role of the anterior insula and basal ganglia in judging an act to be morally wrong likely represents a general role for these regions in encoding negative valence and avoiding aversive stimuli rather than a unique role in contributing to negative moral verdicts. (Borg et al. 2011, 408)

Across the board, these researchers are attributing the process of moral judgment formation to parts of the brain that play more general judgment formation roles. In short, these brain processes are not exclusively reserved for what we call 'moral' judgments.

Perhaps more significantly, a neuroimaging study by Borg and colleagues (2011) found evidence that the brain may rely on separate neural systems for reaching positive and negative 'moral' verdicts (2011, 409). If there are separate systems for reaching positive and negative 'moral' judgments,

then how do we reconcile this with the assumption that moral judgments are the result of a unified process about which we can make meaningful generalizations?

There is additional evidence for the disunity of 'moral' judgment at the level of brain processes. Liane Young and James Dungan conducted a review of neuroimaging research on moral judgment and concluded that "morality [relies] on domain general-processes which are housed in many parts of the brain (…) morality is virtually everywhere in the brain" (2012, 1).

Again, this conclusion has been reached by other researchers, such as John Decety and Jason Cowell, when they write:

> What has become clear from social and clinical neuroscience research is that there is no unique center in the brain for moral judgment. Rather there are interconnected systems that are not domain specific but support more domain-general processing, such as affective arousal, attention, intention, understanding, and decision making. (Decety and Cowell 2014, 528-529)

At this point, the cognitive science community has coalesced around the following conclusion: there is no single area/system of the brain, nor any stable and clear process patterns, that play the primary judgment formation role when people make judgments about what we typically classify as 'moral' issues. They now believe that the formation of moral judgments is dependent on a variety of areas of the brain, and these areas are not primarily devoted to moral judgment; that is, what we call 'moral' judgment is derived from functional areas of the brain primarily devoted to other tasks.[8]

This seems to imply, then, that the process of making a moral judgment likely shares features with what we previously would have taken to be other, distinct judgment types. Thus, if moral judgments are derived from functional areas of the brain that are primarily devoted to other tasks, then it is unclear how this can serve to typologically differentiate 'moral' judgments from judgment simpliciter or other types or judgments.

If we step back a moment and consider this approach to understanding moral judgment, we can see that it was likely to yield such unsatisfying

---

[8] See also Sackris and Larsen (2022) and Sackris (2022) for overviews of neuroscientific research on moral judgment formation. They reach a similar conclusion to the one formed here: at this time, moral judgment formation has not been reliably correlated with any distinctive brain process, set of processes, or set of brain areas.

results. As we stated at the beginning of this paper, if we were to ask ourselves whether we think that *judgments* are made by a single area of the brain or whether *judgment* called on the same set of processes in every context, our *prima facie* answer would most likely be 'No'. The term 'judgment' is used to capture the result of an array of psychological processes: from almost instantaneous belief formation (e.g., when a person judges that another person is attractive) to the result of a careful deliberation process (when a person offers the solution to a complex problem in mathematics). Surely these disparate psychological processes must correspond to different brain processes or systems. If it seems unlikely that 'judgment' itself could be successfully identified with a distinct set of brain processes/single region of the brain using neuro-imaging technology, then given the variety of contexts that give rise to 'moral' judgments, we also have little reason to think that 'moral' judgment could be successfully identified with a single distinctive brain area or set of processes that are easily distinguished from other judgment types.

4.   Defining Moral Judgments on the Basis of Conceptual and Introspective Features

In this section, we consider whether moral judgments can be typologically defined through conceptual and introspective analysis. That is, even if moral judgment is not unified by its *content* or at the level of *brain processes*, perhaps the way we conceive of and conceptualize moral judgments unifies them as a distinctive type or kind. Afterall, diverse processes could still yield similar results in specific, important respects. For example, we may always conceive of moral judgments as having certain, necessary features that other judgment types simply do not have. On such a view, it is not the content being judged, nor the brain processes engaged that make a judgment a 'moral' one, but instead the way the judger reflectively thinks about their judgment that differentiates it from other judgment types.

Such a view was put forward by R.M. Hare (1981, 53-56), as he defined 'moral judgments' as judgments that are conceived as: universalizable, prescriptive, and overriding (for a criticism, see Sinnott-Armstrong and Wheatley 2012). According to Hare, the key feature of the term 'moral' is its 'overriding-ness'. We make a great many prescriptions in our ordinary lives; that is, there are various uses of the word 'ought' and 'must'. However, when two prescriptions conflict, the one that is conceived as being 'moral' will take precedence over the other prescription (1981, 55-56).

A variant of Hare's approach has gained some appeal among contemporary philosophers, where it has recently been advanced by Victor Kumar in a series of papers (2015; 2016a; 2016b). Adding to Hare's definition, Kumar claims that speakers conceptualize moral judgments as (1) serious, (2) general, (3) authority-independent, and (4) objective (e.g., Kumar 2015). By "serious", Kumar means, like Hare, that moral judgments are "overriding" in relation to other judgments. By "general", Kumar means that moral judgments are conceived of as not bound by place and time. By "authority-independent", Kumar means that people conceive of moral/immoral acts as being right/wrong even in circumstances when some authority says that the act is permissible/impermissible to commit. By "objective", Kumar means that moral judgments are conceived as akin to judgments about matters of fact. Overall, Kumar argues for the position that moral judgments constitute a distinctive type of judgment (he argues that they are a natural kind) based on his claim that they share this cluster of four features that distinguish them from other judgment types.

As Kumar is one of the few contemporary authors who explicitly follows Hare's approach and argues for the view that moral judgments constitute a distinctive kind or type, we will carefully consider his position.

On Kumar's view, the four features outlined above form a "homeostatic property cluster" and he proposes that "the human cognitive system is organized in such a way that the four features have a nomological tendency to cluster together" (2015, 2896). That is, these four features co-occur in a law-like manner and distinguish moral judgments from other kinds of judgments (2015, 2889-2890). Kumar maintains that these shared features indicate that moral judgment is, in fact, a unified phenomenon and the presence of these four features is what distinguishes moral judgments from other judgment types.

We might ask why Kumar settles on exactly these four features (and thereby neglects others)? According to Kumar, the four features can be derived from analyzing research on the so-called *moral/conventional* distinction. This body of research has been interpreted to show that individuals (including children as young as three years old) can reliably distinguish between two fundamentally different forms of (normative) violations, namely, moral and conventional violations (cf., Gilligan 2016; Witherell and Edwards 1991). As reported in a handful of studies, when participants are asked to explain the basis for this distinction, some will readily classify moral violations as more wrong than conventional violations; they see moral violations as more authority-independent; they see moral violations as time and place independent; and they explain the wrongness of moral violations in terms of the harm they cause to others,

which they do not typically do for conventional violations (e.g., Nucci and Turiel 1978; Smetana 1981; Turiel 1983; for a recent review see Margoni and Surian 2020).

It is this routine distinction between conventional violations and moral violations, and the rationales that study participants typically give for such a distinction, that leads Kumar to conclude that it is features (1)-(4) that mark out moral judgments as distinctive.

Although Kumar's approach appears to be based on careful, empirical observations, it nevertheless faces the problem of being open to clear counterexamples. Indeed, for Kumar's definition to fail, all that needs to be shown is that there are some clear cases of 'non-moral' judgments that have the four properties Kumar has identified; or that there are judgments typically classified as moral that lack one (or more) of the four features identified by Kumar.

Evidence of such counterexamples are abundant in the literature. For instance, a study by Daniel Kelly and colleagues (2007) shows that subjects are reliably willing to view their moral judgments as non-objective when considering certain moral vignettes. Similarly, survey-studies by James Beebe and David Sackris (2016) and Geoffrey Goodwin and John Darley (2008) have shown that people do not necessarily conceive of morality as objective. [9] To be more specific, empirical work has consistently demonstrated that individuals do not always conceive of moral judgments as objective or generalizable. Therefore, Kumar's claim that moral judgments share the universal feature of objectivity can be refuted on empirical grounds.

When confronted with such strong evidence, Kumar does not disregard the validity of the data, but states instead that the "findings may show that morality is not conceived as universal, but they do not show that morality is conceived as no more general than convention" (2015, 2899). Kumar continues this line of defense when he goes on to say, "a few deliberately chosen cases in which the components come apart is quite a long way from disconfirming evidence" (2015, 2899). Finally, Kumar also states that his theory "cannot be refuted simply by pointing to cases of moral judgment in which *one* of the features are absent" (2015, 2901, our italics).

---

[9] Kumar admits in footnote 5 of his (2015) that Goodwin and Darley have shown that subjects do not treat all moral judgments as objective, however he doesn't take this as evidence against his view, as he attributes it to study participants doubting that such disputes could be rationally adjudicated. He does not discuss Beebe and Sackris. A recent paper by Paul Rehren and Sinnott-Armstrong (2022) indicates that individuals are quite willing to change their mind regarding earlier moral judgments, which also might lead us to question attributions of objectivity.

What, then, would count as disconfirming/falsifying evidence of Kumar's definition of moral judgment? He says that "if clearer counterexamples could be produced, in which judgments that are intuitively classified as moral lack several of the four features, that would count against the view" (2015, 2903). That is, Kumar seems to believe that finding a judgment that is clearly conceived of as moral, yet lacks two of the four features, is the only legitimate way to refute his view.

Although we already see his view as defeated by the empirical evidence discussed and we fail to see why anyone should accept such a high bar for the refutation of Kumar's definition that he sets, we think that it is indeed possible to accommodate Kumar's challenge. Admittedly, it is difficult to conceive of a paradigm moral judgment that is widely taken to lack two of the four identified features. Part of the reason for this is, as we discussed above, that there is general disagreement as to what counts as a moral judgment. However, we will give it a try. Here is what we consider to be a moral judgment that is both *not* serious and *not* authority-independent:

> You shouldn't smoke

If we tell someone that they shouldn't smoke, unless we are ourselves medical doctors or spent our time reading medical journals, our judgment is almost entirely authority *dependent*; in fact, someone's advice against smoking might be based entirely on the Surgeon General's warning placed prominently on the pack of cigarettes. Such a person might also be willing to change their judgment if the Surgeon General changed their advice. Second, many people evidently do not take such judgments to be all that serious, that is, these judgments don't necessarily override all other interests, such as their own or the person they might tell this to: most people likely won't knock the cigarette out of someone else's mouth or crush up their entire pack, and probably doesn't expect the person in question to quit that very instant.

Although smoking primarily effects the person who engages in it, it does affect others via second-hand smoke and by setting a bad example, say, to impressionable children. However, even though said behavior has the potential to negatively impact others, we do not actively view such behavior as a *serious* matter that must be *immediately* attended to. For example, a smoker might say to themselves "I really need to stop smoking in front of my children" but then later in the day go ahead and do so anyways for a variety of reasons.[10]

---

[10] 'You should drive an electric car' might be another example of a moral prescription that is both not serious and authority-dependent.

If the smoking-example is unconvincing to some readers, they might consider the following religious injunctions to be examples of moral judgments that are based on authority and not considered all that serious or over-riding by the practitioners. For example, a practitioner of Mormonism is not supposed to consume "hot drinks" on the authority of Joseph Smith but may do so on occasion as they don't view the injunction as all that serious. On further consideration, one might claim that many human beings base some of their moral judgments on some version of a religious framework, which seems to imply that such judgments by definition are (1) authority-dependent (e.g., the written dogmas), (2) non-objective (e.g., amenable by higher institutional authority), (3) non-general (e.g., rules only apply to the religious cohort), and (4) non-serious (e.g., dogmas will not necessarily overrule other judgments).

We may also refute Hare and Kumar's approach by identifying judgments that are not typically conceived of as moral yet have the properties of being (1) serious, (2) general, (3) authority independent, and (4) objective. That is, there are seemingly clear examples of non-moral judgments that have all four features identified by Kumar.

For example, consider a philosopher who, after reviewing a great number of arguments, makes the judgment that moral sentimentalism is true—and not moral cognitivism—such a judgment is perhaps regarded by said philosopher as serious in that it overrides some of his earlier judgments about the nature of morality, general, authority-independent, and objective. Does this make it a moral judgment?

This example demonstrates what we see as perhaps the biggest challenge for Kumar's definition, namely, that it is too *inclusive*. Indeed, there is an endless array of judgments that have all four features that Kumar outlines, yet are by no means traditionally considered to be moral judgments. We imagine that some art critics take their judgments to be serious, general, authority-independent, and objective. We suppose that most scientific judgments have these properties as well. When Galileo judged that the Earth circum-navigated the sun, this was serious (it overrode other judgments), general, authority-independent, and objective. Are all these judgments also moral judgments? On Kumar's view, they most certainly are. And perhaps for that reason alone, Kumar's definition falls apart.

While Kumar's definition of moral judgment seems to admit of counterexamples (e.g., as being empirically falsified or too inclusive), we could also raise a concern about the evidence it rests upon. First, the significance of the moral-conventional distinction has for many years been disputed (e.g., Witherell and Edwards 1991). For example, some

researchers have demonstrated that the way a question is framed has a significant influence as to whether study participants construe the violation of a norm as moral or conventional (e.g., Margoni and Surian 2020).

It may be that in some cases moral judgments are conceived as more serious, or more objective than other judgments; however, in other contexts it may be that moral judgments are not considered all that serious or taken to transcend a given culture. The results from survey-studies (e.g., Beebe and Sackris, 2016; Goodwin and Darley, 2008) indicate that the relationship between moral and conventional violations is much more akin to two different end points on a *single* scale instead of two distinct *kinds* of normative violations. If we asked four-year-old children whether they should ever touch a hot stove, they would likely say "No" regardless of whether an authority figure told them it was okay to do so or even if they were in a foreign setting. On Kumar's account, this would make prohibitions against touching hot stoves seem to fall into the category of moral injunctions (especially since on his own account not all four features have to be present). If this seems like an inappropriate categorization, then his definition of "moral judgment" has likely missed its mark.

5.   Concluding Remarks

In this paper we have argued that the long-standing philosophical assumption that moral judgments are typologically distinct from other judgment types should be questioned, and possibly rejected. In an Aristotelian framework where definitions adhere to the logical constraints —A is a B that Cs—philosophers and scientists have yet to offer clear definitions of the 'Cs'. Further supporting the conclusion that moral judgments are not typologically discrete is the observation that philosophers often omit defining what a *judgment* is. Logically, it is impossible to build a sufficiently meaningful taxonomy without defining the genus (i.e., the 'B') that we are attempting to place species under (i.e., the 'A' that 'Cs').

As contended throughout this manuscript, although philosophers rarely say this explicitly, they seem to sometimes suppose that moral judgments are always arrived at via some single uniform process. From such a position, it is intuitive to stipulate that there is a 'moral judgment' area in the brain, or that moral judgments must *always* have certain processes or features. However, neuroscientific evidence tells a different story, that judgments generally are formed in a variety of ways. The same seems to be true of moral judgments. Some moral judgments are arrived at almost instantaneously; other moral judgments are only arrived at as a result of

deep reflection. Admittedly, the products of different processes can still have similar properties. However, if we recognize that moral judgments can be arrived at in these seemingly very different ways, why are we so tempted to imagine that they must have shared, necessary features?

If we can use the word 'judgment' to describe the result of highly diverse cognitive processes that range from instantaneous belief formation to the production of a proof in propositional logic, we have to consider the possibility that what we call 'moral judgments' can typically be the result of highly diverse cognitive processes as well; we should also entertain the possibility that these processes are capable of yielding differing results: perhaps in some cases beliefs, in some cases desires, in some cases some third mental state, such as *besires*.[11] This doesn't mean we should throw out the term 'moral judgment'; but it does mean that we may need to reconceive our inquiry into judgments that we typically refer to as 'moral' and recognize that few sweeping generalizations about such judgments are likely to be true.

Acknowledgments

REFERENCES

Altham, James. 1986. "The Legacy of Emotivism." In *Facts, Science, and Morality: Essays on A.J. Ayer's Language, Truth and Logic*, edited by Graham McDonald and Crispin Wright, 275-88. Oxford: Blackwell.

Arp, Robert, Barry Smith, and Andrew Spear. 2015. *Building Ontologies with Basic Formal Ontology*. Cambridge, MA: MIT Press.

Berg, Jan. 1982. "Aristotle's Theory of Definition." *Atti del convegno internazionale di storie della logica*, 19-30.

Borg, Jana, Walter Sinnott-Armstrong, Vince Calhoun, and Kent Kiehl. 2011. "Neural Basis of Moral Verdict and Moral Deliberation." *Social Neuroscience* 6 (4): 398–413. https://doi.org/10.1080/17470919.2011.559363

Brzović, Zdenka. 2018. "Natural Kinds." In *Internet Encyclopedia of Philosophy*. Accessed May 2, 2023. https://www.iep.utm.edu/nat-kind/

---

[11] According to Smith, this term was coined by J.E.J. Altham. For a discussion of the term, see Smith (1994, 118-125).

Campbell, Richmond. 2007. "What is Moral Judgment?" *The Journal of Philosophy* 104 (7): 321-49.

Carroll, Noël. 2012. "Recent Approaches to Aesthetic Experience." *Journal of Aesthetics and Art Criticism* 70 (2): 165-77.

Carroll, Noël. 2015. "Defending the Content Approach to Aesthetic Experience." *Metaphilosophy* 46 (2): 171-88. https://doi.org/10.1111/meta.12131

Cushman, Fiery and Liane Young. 2011. "Patterns of Moral Judgment Derive from Nonmoral Psychological Representations." *Cognitive Science* 35 (6): 1052-75. https://doi: 10.1111/j.1551-6709.2010.01167.x

Dancy, Jonathan. 2018. Moral Particularism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed June 15, 2022. https://plato.stanford.edu/archives/win2017/entries/moral-particularism/.

Decety, Jean and Thalia Wheatley, eds. 2015. *The Moral Brain: A Multidisciplinary Perspective*. Cambridge, MA: MIT Press.

Decety, Jean, and Jason Cowell. 2014. "Friends or Foes: Is Empathy Necessary for Moral Behavior?" *Perspectives on Psychological Science* 9 (4): 525-37. https://doi/10.1177/1745691614545130

Dreier, James. 1996. "Accepting Agent Centred Norms: A Problem for Non-Cognitivists and a Suggestion for Solving it." *Australasian Journal of Philosophy* 74 (3): 409–22.

Fiske, Alan, and Tage Rai. 2014. *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor*. Cambridge: Cambridge University Press.

Flanagan, Owen. 1993. *Varieties of Moral Personality*. Cambridge, MA: Harvard University Press.

Foot, Philippa. 1967. "Abortion and the Doctrine of Double Effect." *Oxford Review* 5: 5-15.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge. MA: Harvard University Press.

Gilligan, Carol. 2016. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.

Glasgow, Joshua. 2013. "Does Direct Moral Judgment Have a Phenomenal Essence?" *Journal of Moral Philosophy* 10 (1): 52-69. https://doi.org/10.1163/174552412X628841

Goodwin, Geoffrey, and John Darley. 2008. "The Psychology of Meta-ethics: Exploring Objectivism." *Cognition* 106 (3): 1339-66. https://doi.org/10.1016/j.cognition.2007.06.007

Greene, Joshua, R. Brian Sommerville, Leigh Nystrom, John Darley, and Jonathan Cohen. 2001. "An fMRI Investigation of Emotional

Engagement in Moral Judgment." *Science* 293 (5537): 2105-08. https://doi:10.1126/science.1062872

Greene, Joshua, Leigh Nystrom, Andrew Engell, John Darley, and Jonathan Cohen. 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44 (2): 389-400. https://doi:10.1016/j.neuron.2004.09.027

Greene, Joshua. 2008. "*The Secret Joke of Kant's Soul.*" In *Moral psychology, Vol. 3. The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott Armstrong, 35–80. Cambridge, MA: MIT Press.

Greene, Joshua. 2013. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*: Penguin Press.

Greene, Joshua. 2015a. "The Cognitive Neuroscience of Moral Judgment and Decision-making." In *The Moral Brain: A Multidisciplinary Approach*, edited by Jean Decety and Thalia Wheatley, 197-220. Cambridge, MA: MIT Press.

Greene, Joshua. 2015b. "The Rise of Moral Cognition. *Cognition* 135: 39-42. https://doi.org/10.1016/j.cognition.2014.11.018

Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Pantheon Books.

Hare, Richard. 1981. *Moral Thinking: Its Levels, Methods, and Point*. New York: Oxford University Press.

Kant, Immanuel. 1998. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press.

Kelly, Daniel, Stephen Stich, Kevin Haley, Serena Eng and Daniel Fessler. 2007. "Harm, Affect, and the Moral/Conventional Distinction." *Mind and Language* 22 (2): 117-31. https://doi.org/10.1111/j.1468-0017.2007.00302.x

Kumar, Victor. 2015. "Moral Judgment as a Natural Kind." *Philosophical Studies* 172 (11): 2887-2910. https://doi.org/10.1007/s11098-015-0448-7

Kumar, Victor. 2016a. "Psychopathy and Internalism." *Canadian Journal of Philosophy* 46 (3): 318-45. https://doi.org/10.1080/00455091.2016.1165571

Kumar, Victor. 2016b. "The Empirical Identity of Moral Judgment." *The Philosophical Quarterly* 66 (265): 783-804. https://doi.org/10.1093/pq/pqw019

Marek, Scott, Brenden Tervo-Clemmens, Finnegan Calabro, et al. 2022. "Reproducible Brain-wide Association Studies Require Thousands of Individuals." *Nature* 603 (7902): 654-60. https://doi.org/10.1038/s41586-022-04492-9

Margoni, Francesco, and Luca Surian. 2020. "Question Framing Effects and the Processing of the Moral-conventional Distinction." *Philosophical Psychology* 34 (1): 76-101. https://doi.org/10.1080/09515089.2020.1845311

Martin, Wayne. 2006. *Theories of Judgment: Psychology, Logic, Phenomenology*. Cambridge: Cambridge University Press.

May, Joshua and Victor Kumar. 2018. Moral Reasoning and Emotion. In *The Routledge Handbook of Moral Epistemology*, edited by Karen Jones, Mark Timmons, and Aaron Zimmerman, 139-56. London: Routledge.

Moore, George Edward. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Nucci, Larry, and Elliot Turiel. 1978. "Social Interactions and the Development of Social Concepts in Preschool Age Children." *Child Development* 49 (2): 400-07. https://doi.org/10.2307/1128704

Osterloff, Emily. Nd. Brontosaurus: Reinstating a prehistoric icon. *Natural History Museum*. Accessed January 24, 2023. https://www.nhm.ac.uk/discover/brontosaurus-reinstating-a-prehistoric-icon.html#:~:text=That%20was%20until%20a%20study,status%20of%20this%20iconic%20dinosaur

Prinz, Jesse. 2007. *The Emotional Construction of Morals*. New York: Oxford University Press.

Rehren, Paul and Walter Sinnott-Armstrong. 2022. "How Stable are Moral Judgments? *Review of Philosophy and Psychology*, https://doi.org/10.1007/s13164-022-00649-7

Sackris, David. 2022. "The Disunity of Moral Judgment: Implications for the Study of Psychopathy." *Philosophical Psychology*. https://doi.org/10.1080/09515089.2022.2155125

Sackris, David, and Rasmus Rosenberg Larsen. 2020. "A Consideration of Carroll's Content Theory." *Journal of Value Inquiry*, 54: 245-55. https://doi.org/10.1007/s10790-019-09693-6

Sackris, David, and Rasmus Rosenberg Larsen. 2022. "The Disunity of Moral Judgment: Evidence and Implications." *Philosophical Psychology*. https://doi.org/10.1080/09515089.2022.2056437

Sackris, David, and Rasmus Rosenberg Larsen. (2023). "Are There 'Aesthetic' Judgments?" *Erkenntnis*. https://doi.org/10.1007/s10670-023-00663-w

Sayre-McCord, Geoff. 2014. Metaethics. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed June 15, 2022. https://plato.stanford.edu/archives/sum2014/entries/metaethics/.

Shafer-Landau, Russ. 2015. *The Fundamentals of Ethics*. Oxford: Oxford University Press.

Seppälä, Selja, Alan Ruttenberg, and Barry Smith. 2017. "Guidelines for Writing Definitions in Ontologies. *Ciência Da Informação* 46 (1): 73-88.

Sinnott-Armstrong, Walter, and Thalia Wheatley. 2012. "The Disunity of Moral Judgment and Why it Matters in Philosophy." *The Monist* 95 (3): 355-77. https://doi.org/10.5840/monist201295319

Sinnott-Armstrong, Walter, and Thalia Wheatley. 2014. "Are Moral Judgments Unified?" *Philosophical Psychology* 27 (4) 451-74. https://doi.org/10.1080/09515089.2012.736075

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell Publishing.

Stich, Steven. 2006. "Is morality an elegant machine or a kludge?" *Journal of Cognition and Culture* 6 (1-2): 181-89. https://doi.org/10.1163/156853706776931349

Thomson, Judith. 1976. "Killing, Letting Die, and the Trolley Problem." *The Monist* 59 (2): 204-17. https://doi.org/10.5840/monist197659224

Verplaetse, Jan. 2009. *Localizing the Moral Sense: Neuroscience and the Search for the Cerebral Seat of Morality., 1800-1930.* New York: Springer. https://doi:10.1007/978-1-4020-6322-0

Young, Liane, and James Dungan. 2012. "Where in the Brain is Morality? Everywhere and Maybe Nowhere." *Social Neuroscience* 7 (1): 1-10. https://doi:10.1080/17470919.2011.569146

Witherell, Carol, and Carolyn Edwards. 1991. "Moral Versus Social-conventional Reasoning: A Narrative and Cultural Critique." *Journal of Moral Education* 20 (3): 293-304. https://doi:10.1080/0305724910200306

FREE WILL AS AN EPISTEMICALLY INNOCENT FALSE BELIEF

Fabio Tollon[1,2,3]

[1] Department of Philosophy, Bielefeld University, Germany
[2] Department of Philosophy, Center for Applied Ethics, Stellenbosch University, South Africa.
[3] Centre for Artificial Intelligence Research (CAIR), Pretoria, South Africa.

ABSTRACT

In this paper I aim to establish that our belief in free will is epistemically innocent. Many contemporary accounts that deal with the potential "illusion" of freedom seek to describe the pragmatic benefits of belief in free will, such as how it facilitates or grounds our notions of moral responsibility or basic desert. While these proposals have their place (and use), I will not explicitly engage with them. I aim to establish that our false belief in free will is an epistemically innocent belief. I will endeavour to show that if we carefully consider the circumstances in which particular beliefs (such as our belief in free will) are adopted, we can come to better appreciate not just their psychological but also their epistemic benefits. The implications, therefore, for future investigations into the philosophy of free will are that we should consider whether we have been too narrow in our pragmatic defences of free will, and that we should also be sensitive to epistemic considerations.

Keywords*: free will; epistemic innocence.*

(A2)1

Introduction

In this paper I will argue that belief in free will is epistemically innocent. Some authors have argued that even if belief in free will is false, it might be recommended on pragmatic grounds. I suggest another reason why belief in free will might be good. More specifically, I argue that our belief in free will has certain, otherwise unavailable, epistemic benefits. To do so I rely on work done on the epistemic status of beliefs, developed by Lisa Bortolotti and her research team (2020). Of course, if such a belief about free will is *true*, this obviously makes it a good belief to have. My argument is therefore that *even if* this belief turns out to be false, it is still a good belief to have, for hitherto unappreciated reasons.[1] An important upshot of this account, therefore, is that it provides a novel mechanism for exonerating free will beliefs (if they turn out to be false).

Many contemporary accounts which deal with the "illusion" of freedom seek to describe the pragmatic benefits of belief in free will, such as how it facilitates or grounds our notions of moral responsibility or basic desert[2] (Mele 2005; Wegner 2002; Smilansky 2000; Strawson 2010). While these proposals have their place (and use), I will not explicitly engage with them. I aim to establish that our (potentially false) belief in free will is an epistemically innocent one, and that this holds independently of whether we do in fact have free will or not.

Human beings have long been considered the prime example of rationality. However, the empirical literature suggests that we are not as rational as we were traditionally conceived to be (Bortolotti 2015a, 1). This is not to claim that we are at base irrational or insane, but rather to point out that our beliefs are not *always* guided by reason, and that biases, heuristics, and affect all come to play a role in how we reason. In this paper, therefore, I aim to bring two distinct research projects together for the first time: The literature on the epistemic innocence of beliefs, and the literature on free will.

One of the main meta-justifications for this epistemically-orientated approach is one that informs most of philosophy: A desire to get, at the very least, closer to the truth. While pragmatically belief in free will certainly provides benefits, the epistemic benefits I consider in this paper provide us with an additional, distinct set of reasons for evaluating our

---

[1] My thanks to an anonymous reviewer for this suggested framing.
[2] See Sommers (2007, 64) for an evolutionary argument for how it is in fact the other way around: the belief in robust moral responsibility leads to the belief in free will.

belief in free will. In order to show that this is the case regarding belief in free will, I will move beyond purely pragmatic concerns and show how this false belief can lead to epistemic gains for agents who adopt it.[3]

I will proceed as follows. First, I put forward that I assume some version of 'illusionism' about free will. Second, I outline what it means for a belief to be *epistemically innocent*. Third, I apply this to belief in free will, and show what unique, otherwise unobtainable, epistemic benefits it might accord.

1.   The Way Forward

We might not have free will. If this is true, it would be false for us to believe that we have free will. We find support for what has been termed 'illusionism' in the philosophical literature on free will (Smilansky 2000; Strawson 2010). Smilansky, for example, explicitly endorses such a view of free will. He uses the term 'illusion' because it draws our attention to the "various ways in which false beliefs are held, without complete awareness of their falseness, in the face of stronger epistemic claims to the contrary" (Smilansky 2000, 148). For my purposes, and in what follows, I will bracket the question of what exactly free will is, and whether it really is an 'illusion' in the sense above. Instead, I will proceed *as if* our belief in free will is simply false, and then show how it might be exonerated on epistemic grounds.

There are many ways to understand what 'free will' might mean: Is it alternative possibilities for choice and action, freedom from causal determination, the ability to act or refrain from acting at a certain time, or the ability for an agent to act rationally? I cannot settle this debate here, but from these different senses of what free will means, we can get a rough sense of what a 'belief' in free will might entail. Such a belief, from the perspective of the agent who adopts it, would have something to do with control, knowledge, and action. In what follows, therefore, it is these general aspects of free will that I focus on in my articulation of the potential epistemic benefits.

So, assuming that our belief in free will is false, one way to exonerate such a faulty belief is by appealing to *pragmatic* upshots of the belief. The

---

[3] Pragmatic considerations are traditionally focused on the useful consequences of a specific belief (i.e., things going *well*). The scope of pragmatic evaluation is therefore quite large and includes things like the psychological or broader societal benefits/costs of the cognition in question. Epistemic considerations are much narrower in their scope, and hone in on what implications certain beliefs could have on our ability to acquire knowledge about the world (getting things *right*).

broadly pragmatic benefits of this belief are relatively clear, such as facilitating our ascriptions of moral responsibility, allowing us to view ourselves as 'in control' of our actions, etc. The question I am interested in, however, is whether there are also epistemic benefits to this type of belief. While in an ideal world our beliefs would all be supported by and responsive to the available evidence, the limited cognitive capacities we have as agents ultimately leads to us adhering to some poorly supported beliefs, which may be unresponsive to evidence. On the face of it, one might think it rational to dismiss these epistemically dubious claims altogether. Yet in some cases a *prima facie* epistemically costly cognition[4] can in fact lead to positive epistemic outcomes (Bortolotti 2015b). I will argue that this is the case regarding free will. An important upshot of this account is that it provides further reasons for exonerating belief in free will.

2.   Epistemic Innocence

In this section I will show that even if our belief in free will is false, it is nonetheless an epistemically innocent belief. This argument steers clear of the metaphysical problems introduced earlier and simultaneously moves beyond the traditional pragmatic focus of other theories in such debates (Dennett 1984). I will endeavour to show that if we carefully consider the circumstances in which particular beliefs (such as our belief in free will) are adopted, we can come to better appreciate not just their *psychological* but also their *epistemic* benefits. This strategic approach to our beliefs allows us to guard against the 'trade-off' view regarding certain types of false beliefs.

The trade-off view assumes that while certain faulty cognitions may offer psychological or pragmatic benefits, these benefits come with epistemic costs (Letheby 2016, 31). To put it differently, the trade-off view assumes that there are *only* epistemic costs associated with faulty beliefs, and no associated benefits. These costs are presumed to stem from the irrational nature of the belief, as it might be unresponsive to evidence, implausible and/or not an accurate representation of reality (Bortolotti 2015, 492). However, this trade-off view presents us with an overly simplistic representation of what is going on regarding both the formation and retention of our beliefs.

---

[4] An epistemically costly cognition is one in which certain epistemically healthy norms are violated, such as when intentions do not match beliefs and desires, when goals are not pursued consistently, or when beliefs are badly supported by evidence and conflict with the science of the day (Bortolotti 2015b, 3). As philosophers, our meta-commitment to uncovering the truth means that epistemic criteria generally trump pragmatic ones.

In many cases there is considerable overlap between both pragmatic and epistemic criteria, as in situations when psychological well-being is positively correlated with increased social engagement, leading to the formation of more true beliefs over time. However, there are also cases in which epistemic and pragmatic considerations can compete with one another, such as in cases where irrational beliefs can be pragmatically beneficial,[5] or when rational beliefs are not useful.[6] For a belief to be innocent is to suggest that even though it is epistemically irrational, it might nonetheless confer certain benefits which could act as an *excuse* for holding the belief. Thus, the notion of 'innocence' at work here is an application of the "sense of innocence as absence of guilt to the epistemic domain" (Bortolotti 2020, 9).

The type of agents we are necessarily implies that we are limited by certain physical constraints in our ability to coherently form and maintain our beliefs. Our limited cognitive capacity often leads us to adopt poorly supported beliefs, which often act as helpful heuristics as opposed to facilitating proper reasoning. Kahneman (2011) gives the example of what he calls the "affect" heuristic.[7] The affect heuristic is a cognitive shortcut which allows agents to efficiently solve problems by relying on their *current mood*. It allows people to judge the risk or benefits of a specific action by relying on which *feelings* are associated with that outcome, as opposed to engaging in temporally expensive reasoning. There are cases where this can be useful (better avoid this spider) or misleading (climate change does not produce an affective response in many, and so is thought by some to not be a serious issue).

While it might be reasonable to dismiss these types of 'epistemically costly' cognitions[8] altogether, there are times when an epistemically costly cognition can enhance our long-term epistemic functionality, such as our ability to form more true beliefs over time (Bortolotti 2016, 888). A classic example cited in the literature is that of BX, a former musician who, after

---

[5] For example, imagine that someone believes that distant celestial bodies have a meaningful causal impact on the unfolding of their lives. Based on this they decide to make a drastic change in their lifestyle (such as adjusting their eating habits or purchasing a specific type of coloured rock), which leads to positive, practical consequences. This belief is clearly absurd (and, sadly, widespread) but can lead to positive outcomes.

[6] For example, imagine someone who believes that it is their job to *constantly* tell the truth, no matter what. This also involves telling their partner that yes, they do look bad in those jeans. Such a person is unlikely to have many close friendships. While they may be right, they lack the social nuance sometimes required to generate and sustain meaningful interactions with others.

[7] See Kahneman (2011, 101-175) for a detailed and practical account of the precarious nature of our so-called "reasoning" capabilities. Kahneman shows how our thinking is heavily influenced by cognitive heuristics that allow us to reason faster, but not necessarily better. Examples of these identified by Kahneman include the mood heuristic, the affect heuristic, and the availability heuristic, to name a few.

[8] Cognitions that violate healthy epistemic norms, such as being unresponsive to evidence, etc.

a car accident, became a quadriplegic (Bortolotti 2015b, 492). Before the accident BX was in a healthy relationship, but soon after the incident his partner broke up with him. Following this BX developed the delusional belief that his partner was still with him (known as "reverse-Othello syndrome"). This false belief might have allowed BX to get through the trauma of the accident, and so there is a case to be made that it provided psychological benefits. But there are also epistemic benefits to this false belief, such as BX being more willing and able to engage with his doctors and therefore acquire knowledge about how he might best go about the world post-accident. Bortolotti (2015, 495) argues that these cognitions can be construed as being epistemically innocent.

2.1   Epistemic Status

There are two criteria which are necessary and jointly sufficient for a delusional belief to qualify as epistemically innocent:
(1) Epistemic Benefit: The delusional belief confers a significant epistemic benefit to an agent at the time of its adoption.
(2) No Alternatives: Other beliefs that would confer the same benefit are not available to that agent at that time. (Bortolotti 2015, 496).

For the purposes of this paper, however, I am not necessarily endorsing the view that belief in free will is delusional. Rather, I propose more modestly that it might be a false belief. Therefore, with respect to (1) above, we can replace 'delusional' with 'false'.

But just what exactly counts as an epistemic benefit? There are two main lenses one could use when evaluating the epistemic status of a specific belief. Firstly, one could argue that a belief is epistemically advantageous if it allows for the retention or acquisition of true beliefs over time (a veritist). And secondly, one could argue that a belief is epistemically advantageous if it allows for the promotion of an agent's intellectual virtues such as intellectual curiosity or honesty (a virtue epistemologist) (Bortolotti 2016, 889). Taken together these attributes constitute the *epistemic functionality* of an agent, i.e., the ability of the agent to function well epistemically.

In terms of the no alternatives condition, there are three ways in which a cognition may be construed as being unavailable: It may be strictly, motivationally, or explanatorily unavailable (Sullivan-Bissett 2015, 554). A cognition is *strictly* unavailable when it is based on information that is unavailable to an agent via introspection. An example of this would be an agent who suffers from dementia and as a result of which has severe memory impairment. Such an individual may claim to have been at a theme

park in the morning, when the trip actually occurred when they were a teenager. They would be incapable of forming the correct sort of belief regarding the trip because of their memory impairment, and so such a cognition is strictly unavailable. A cognition is *motivationally* unavailable when it is unavailable due to motivational factors. A common example of this type of cognition involves cases of self-deception. Take the case of the cuckolded husband who falsely believes that his wife is faithful to him. He might have evidence that she is being unfaithful, but his desire to believe that she is not having an affair makes the belief that she is unfaithful motivationally unavailable to him. Lastly, a cognition is *explanatorily* unavailable when an agent dismisses it due to its perceived high improbability (Sullivan-Bissett 2015, 554). Consider the example of you finding porcupine quills in your garden. You also observe that there are pieces of your favourite tree missing. It is reasonable at this point to conclude that there is a porcupine chewing the bark off your tree. However, you could also believe that there is a magical fairy that drops porcupine quills and cuts bite-like marks out of trees with a hunting knife. This second type of explanation is dismissed due to its implausibility. It is dismissed because of how unreasonable it seems, and so is explanatorily unavailable when compared to other, more plausible, cognitions.

3. Belief in Free Will as Epistemically Innocent

3.1 Epistemic Benefit

In order for belief in free will to be considered epistemically innocent it must be shown that this belief does in fact provide an epistemic *benefit* to the agent who adopts it. This is not to say that the belief is epistemically *good* overall or free from epistemic *faults*. Rather, it is simply to modestly claim that such faulty cognitions *can* confer *some* epistemic benefits (such as BX being able to continue interacting with his doctor's post-accident and therefore being open to the acquisition of more true beliefs over time) (Sullivan-Bissett 2015, 554). I will show how belief in free will helps us maintain a more coherent sense of self, and, secondly, how it can facilitate the process of reason-giving and taking, which could, for example, help us make various implicit biases explicit (with the hope of their eventual correction). Lastly, I will argue that it heightens our sense of 'perceived control', contributing further to our epistemic well-being.

The first point to consider is the way in which belief in free will facilitates a more coherent sense of self. The mechanism by which this is done turns on the essential causal opacity of certain folk-psychological concepts, such as our beliefs, desires, etc. Beliefs are molar-level phenomena which might

have various correlates at different levels of abstraction (scientific psychology, neuroscience, etc.) (Bortolotti 2010, 2). The implications of this opacity mean that we might be unable to introspect the 'real causes' that lead to the retention or adoption of our beliefs.[9]

However, belief in free will could potentially mitigate this by giving us a plausible 'just so story' about what 'caused' us to act in this or that way. This illusion of competence adds an important sense of *coherence* to our sense of self, which may enhance our self-confidence and well-being (Sullivan-Bissett 2015, 548). With a coherent sense of self an individual is better able to engage with the external world, increasing the possible range of affordances available to them (Bortolotti and Miyazono 2015). It seems reasonable to suppose that individuals who are sure of themselves and feel secure in their beliefs will be more likely to engage with their surroundings.

Furthermore, this type of psychological security might allow for individuals to put themselves in new or perhaps uncomfortable situations, which would grant them new experiences and increase the probability that they acquire more knowledge about the world. An individual is far more likely to be willing to engage with their environment and increase the landscape of affordances available to them if they feel sure of their place in the world.[10] These affordances provide a scheme by which they can expand their knowledge about the world through the sharing of expectations and conventions (Ramstead, Veissiere, and Kirmayer 2016, 4). Such affordances should not be seen as 'things', but rather as *possibilities for action*. These possibilities for action can be viewed as opportunities for an agent to acquire new beliefs about the world, which will also be subject to feedback. In this recursive way an agent may come to gather a significant amount of information about the world around them, leading to an increase in overall *epistemic functionality*. The agent both increases their epistemic virtues by engaging in reason giving and revising certain problematic beliefs and can acquire and retain more true beliefs about the world.

Relatedly, we do not always have direct access to the underlying causes of our actions, and so when we are *questioned* as to why we performed a certain act we often to respond with *post-hoc* rationalizations (Sapolsky 2018, 401). An easy example of this type of explanation (unfortunately) occurs in the self-reports of some explicitly egalitarian individuals. When

---

[9] Furthermore, one can argue that our inherent *a priori* conceptual abilities are not inherently designed for productive introspection, especially when it pertains to comprehending the inner workings of our own minds (see, e.g., McGinn 1989).

[10] An affordance is a possibility for action between an agent and their environment (Ramstead, Veissiere, and Kirmayer 2016, 4).

assessing application documents of students applying for a lab position, faculty members consistently rated male applicants CVs as better suited to the job than their female counterparts (Sullivan-Bissett 2015). The confabulation comes out when the applicants' supporting documents are controlled for: There is no other distinguishing factor between applicants besides their gender, and so it is clear that this gender-bias was causally efficacious in the faculty member's decisions. However, the *reasons given* were that the male candidates were "more competent and hireable" (Sullivan-Bissett 2015). These reasons can be seen as confabulatory as they do not express the 'true' rationale for behaviour.

Therefore, when we engage in the exercise of deliberation, we are almost inevitably engaging in imperfect reasoning strategies, as we cannot have all of the required evidence to make perfect decisions (Bortolotti 2015a, 18).[11] Despite the epistemically faulty nature of our decision-making procedure, however, the *process of reason giving* itself can confer epistemic benefits (Bortolotti 2009). The way this comes about is that through reason giving we might come to better understand ourselves, and, significantly, we may have to make explicit any implicit biases[12] that we have. Once explicit, these *commitments* can be challenged and revised if shown to be false when faced with evidence to the contrary. When conversing with others, we often engage in practices in which we question the intentions that they may have when performing or not performing certain acts ('why don't you donate some of your salary to charity?', 'why do you still eat factory farmed meat?', etc.). Through debate and dialogue with one another we can progressively adopt more epistemically sound beliefs. In order to do this, however, our initial, potentially false, belief that we are free is presupposed. In this way belief in free will, belief that we really are the 'willers' of our actions, opens up the possibility of 'peer-review' for our beliefs.

Additionally, epistemic gains afforded by this potentially faulty cognition are linked to the enhanced sense of psychological well-being associated with "perceived control" (Wegner 2012). People who believe themselves to be the causes of events are more likely to be psychologically healthy. An example of this was uncovered in a study that investigated the coping mechanisms displayed by people who had recently been involved in

---

[11] See Bortolotti (2015a) for a critique of the "rationality assumption" traditionally presupposed in descriptive accounts of human agency.

[12] "Largely unconscious tendencies to automatically associate concepts with one another" (Sullivan-Bissett 2015, 549). Implicit biases are held by most people, even by those who explicitly assert egalitarian positions. See De Houwer et al. (2009) and Nosek et al. (2007) for evidence of this claim. For critical discussion of the implicit biases research program, see, e.g., Machery (2022).

paralyzing accidents (Bulman and Wortman 1977). In brief, the study found that those who attributed the cause of events as external ("someone else did it", "it was random") struggled to cope as well as those who characterized the events internally ("I was responsible"). These feelings of control are therefore positively correlated with psychological well-being, which in turn is correlated with a willingness to engage with one's environment and peers.[13]

We might also think that social interaction encourages us to make explicit our beliefs and facilitates a process of interpersonal hypothesis testing. This interpersonal hypothesis testing is an inherently social phenomenon, insofar as it requires others to listen and potentially respond to what we are saying. It is natural for us to want to be liked and admired by our peers, and so it might be plausible to think that such social interaction might motivate us to do what we think is morally right, increasing our desire to discover something like moral truths. By communicating with others, we become accountable to them and ourselves. We might state our wishes and desires or express our values. Having expressed these there is a pressure to actually follow through: If I claim to be charitable, I had better express this virtue when the situation demands. Of course, there is the very real worry that our social groups might also encourage the formation of epistemically harmful beliefs, but this would not be true in all cases. More importantly for my purposes, however, it is crucial to note that this mechanism enables individuals to gradually acquire a greater number of accurate beliefs, which constitutes an epistemic benefit.

Once again, by making our problematic beliefs explicit we open ourselves up to the opportunity of being proven wrong: Others who might know better than us can correct our faulty beliefs and we can then make an attempt at improving the veracity of our cognitions. In other words, while it might be the case that psychological well-being facilitates this process, there is nonetheless an *epistemic component* to this type of cognition, however minimal it might be. This is not to claim that the epistemic benefits outlined here are stellar. Rather, what I am modestly suggesting is that a simple psychological account of this belief may miss the greater epistemic picture.

3.2   No Alternatives

The second criterion required for a belief to be considered epistemically innocent is the No Alternatives condition. I will suggest that any alternative

---

[13] There is also further evidence that suggests that a sense of control and predictability lower glucocorticoid levels, and therefore reduce stress (Sapolsky 2018, 436).

to the belief in free will is *explanatorily* unavailable (Sullivan-Bissett 2015, 554). It is important to note, however, that even if my argument here is unsuccessful, it does not take away from the epistemic benefits outlined above. Should the No Alternatives condition not hold, belief in free will might not be epistemically *innocent*, but this would not force us to conclude (based on what I have argued above) that it has no epistemic benefits at all. Additionally, the No Alternatives condition refers to specific agents and their beliefs. Thus, it is not a general claim that no belief other than belief in free will is explanatorily unavailable for all agents. Rather, it is about particular agents and their beliefs. In order for this condition to obtain, therefore, it should be the case that from the perspective of the agent no other belief is available that confers the same epistemic benefit.

A belief is explanatorily unavailable when it is "dismissed [by the subject] due to its apparent implausibility" (Sullivan-Bissett 2015, 554). These are cases in which an agent may have certain beliefs about experiences they have had, and where alternative accounts that explain the belief or cognition strike the agent in question as implausible or insufficient (such as the porcupine example introduced earlier). Consider again our feeling of ourselves as unified agents with free will. This feeling, and our subsequent belief in its truth, might be false. However, when we introspect, we are constantly confronted with the fact that we continue to *feel* and *experience* ourselves as free. In other words, no other explanation for the way we feel about the actions we perform is available. Of course, different individuals may report different degrees of freedom: The point is that, for some, such a feeling might be stronger, and so other explanations for their actions may be unavailable.

It is for this reason that I believe that no other belief is explanatorily available. The justification for this claim relies on the fact that there is a close relationship between our perception of each other as free and our ascriptions of moral responsibility. We tend to hold others responsible for their actions because we believe them to be in control of what they are doing, and it is this sense of responsibility that is fundamental to the successful functioning of society. The perception that we are in control of our actions makes us responsible in this morally credible sense.

Before concluding it is important to note a comparison between us believing in free will versus us believing in our lack of freedom.[14] There is difference between us having free will *being true in fact* and us *believing it to be true* (Duus-Otterström 2008, 223). If it is true, this does not mean that people in general or policymakers specifically will radically change

---

[14] My thanks to an anonymous reviewer for pressing me on this point.

the way they act. The real worries arise when we *believe* that we are not free. Now it is beyond the scope of this paper to go into detail with respect to this claim. Suffice to say that it is not at all obvious that us not being free would be that disastrous (Pereboom 2003, 2014). However, I do think there are serious consequences to such a belief that may have a negative effect on overall epistemic functionality. Notions such as basic desert would have no justification, as nobody would ever, strictly speaking, *deserve* anything, as who they are and what they do would not be a product of their *will*. Such responsibility ascriptions are important to the functioning of society more generally, but also for us as individuals. Believing ourselves to be free prompts us to take seriously the fact that we are, in some sense, in control of our actions and can be held accountable for them.

Such accountability leads us to *want to be better*, which causes other epistemic gains, as noted above. We acknowledge that we are responsible for our actions and seek to act in ways which are morally appropriate. This encourages our pursuit of what is truly morally correct, as we strive to improve as moral agents. Being better moral agents implies that we are better *informed* about what to do, and thus have a larger reservoir of information when it comes to making morally laden decisions. It is in this sense that such a belief in free will may have positive epistemic consequences which would not be possible in its absence: we believe we are free, facilitating justified ascriptions of responsibility, which enables us to be better informed moral agents. It therefore seems desirable for us to believe we are free.

4.   Concluding Remarks

In this paper, I have argued that a false belief in free will might be justified on epistemic grounds. There are cases in which it is possible that the adoption of a false belief can prevent an epistemic harm from occurring, and in such a case we may say that the belief is *innocent*. I have argued that free will is one such epistemically innocent belief. This belief, while not epistemically good overall if it turns out to be an illusion, was found to offer clear epistemic benefits to the individual, such as a more coherent sense of self and the acquisition and retention of true beliefs over time.

Furthermore, it was found that no alternatives, other than belief in free will, are explanatorily available which confer the same epistemic benefits. This is perhaps the most obvious weak point of the paper, as there are many people who *do not* in fact believe in free will, and so this belief does indeed seem to be available. Moreover, it seems as though one could not believe

in free will and yet gain the epistemic benefits I've described above. My response to this charge, linked to what I have already said in Section 3.2 above, is that perhaps such a belief is only available to *some agents*. That is, for some, belief in free will is the only means to these epistemic benefits, whereas for others this might not be the case. What exactly might explain this difference is beyond the scope of this paper. Importantly, however, this point is predicted on the fact that the No Alternatives condition is about particular agents and their beliefs, and not about beliefs in general.

To reiterate, this is not to say that this belief is epistemically good overall, but rather to claim that there are at the very least *some* epistemic gains to be had. Consequently, in conjunction with the many pragmatic benefits of this belief that have been the focus in much of the literature (such as an enhanced sense of moral responsibility), there might also be further epistemic benefits that have yet to be explored. The implications, therefore, for future investigations into the philosophy of free will are that we should consider whether we have been too narrow in our pragmatic defences of free will, and that we should also be sensitive to epistemic considerations.

Acknowledgments

REFERENCES

Bortolotti, Lisa. 2010. *Delusions and Other Irrational Beliefs*. New York: Oxford University Press.

Bortolotti, Lisa. 2015a. *Irrationality*. Cambridge: Polity Press. doi: 10.1002/ejoc.201200111.

Bortolotti, Lisa. 2015b. "The epistemic innocence of motivated delusions." Consciousness and Cognition, 33: 490–499. doi: 10.1016/j.concog.2014.10.005.

Bortolotti, Lisa. 2016. "Epistemic benefits of elaborated and systematized delusions in schizophrenia." British Journal for the Philosophy of Science, 67(3): 879–900. doi: 10.1093/bjps/axv024.

Bortolotti, Lisa. 2020. *The Epistemic Innocence of Irrational Beliefs*. New York, NY: Oxford University Press.

Bortolotti, Lisa. and Sullivan-Bissett, E. 2015. "Costs and Benefits of Imperfect Cognitions." *Consciousness and Cognition*, 33: 487–489. doi: 10.1016/j.concog.2015.03.013.

Bortolotti, Lisa. 2009. "The epistemic benefits of reason giving." *Theory and Psychology*, 19 (5): 624–645.

Bortolotti, L. and Miyazono. 2015. "The Ethics of Delusional Belief." *Erkenntnis,* 81(2): 275-296.

Bulman, J. R., and C. B. Wortman. 1977. "Attributions of blame and coping in the 'real world': Severe accident victims react to their lot." *Journal of Personality and Social Psychology* 35: 351–363.

Clarke, R., Capes, J., "Incompatibilist (Nondeterministic) Theories of Free Will". In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Available: https://plato.stanford.edu/archives/spr2017/entries/incompatibilism-theories/ [2018, September 20]

De Houwer, Jan, K. Teige-Mocigemba, A. Spruyt, and A. Moors. 2009. "Implicit Measures: A Normative Analysis and Review." *Psychological Bulletin*, 135 (3): 347–368.

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

Dennett, Daniel C. 2003. *Freedom Evolves*. New York: Viking.

Duus-Otterström, G. 2008. "Betting Against Hard Determinism." *Res Publica,* 14: 219-235.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Great Britain: Allen Lane

Kane, Robert. 1996. *The Significance of Free Will*. Oxford: Oxford University Press.

Letheby, Chris. 2016. "The epistemic innocence of psychedelic states." *Consciousness and Cognition*, 39: 28–37. doi: 10.1016/j.concog.2015.11.012.

Machery, Edouard. 2022. "Anomalies in Implicit Attitudes Research." *WIREs Cognitive Science* 13 (1). https://doi.org/10.1002/wcs.1569.

McGinn, C. 1989. "Can we solve the mind-body problem?" *Mind* 98 (391): 349-366.

Mele, Alfred R. 2005. "Critique of Pereboom's 'Four Case Argument' for Incompatibilism." *Analysis* 65 (1): 75-80.

Nosek, B. A., F. L. Smyth, J. J. Hansen, T. Devos, N. M. Lindner, K. A. Ranganath, et al. 2007. "Pervasiveness and correlates of implicit attitudes and stereotypes." *European Review of Social Psychology*, 18 (1): 36–88.

Pereboom, Derk. 2003. *Living without Free Will*. Cambridge: Cambridge University Press.

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

Ramstead, Max J. D., S. P. L. Veissiere, and L. J. Kirmayer. 2016. "Cultural Affordances: Scaffolding Local Worlds through Shared Intentionality and Regimes of Attention." *Frontiers in Psychology* 7(JUL): 1–21. doi: 10.3389/fpsyg.2016.01090.

Sapolsky, Robert. 2018. *Behave*. London: Vintage.

Sarkissian, Hagop et al. 2010. "Is Belief in Free Will a Cultural Universal?" *Mind and Language* 25 (3): 346–358. doi: 10.1111/j.1468-0017.2010.01393.x

Smilanksy, Saul. 2000. *Free Will and Illusion*. New York, NY: Oxford University Press.

Strawson, Peter F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 187–211.

Strawson, Galen. 2010. *Freedom and Belief*. New York, NY: Oxford University Press

Sullivan-Bissett, Ema. 2015. "Implicit Bias, Confabulation, and Epistemic Innocence." *Consciousness and Cognition* 33: 548–560. doi: 10.1016/j.concog.2014.10.006.

Vihvelin, Kadri. 2017. "Arguments for Incompatibilism." In *The Stanford Encyclopaedia of Philosophy*, edited by E. N. Zalta. Available: https://plato.stanford.edu/archives/fall2017/entries/incompatibilism-arguments/ [2018, September 20].

Wegner, Daniel M. 2002. *Illusion of Conscious Will*. London, England: Bradford Books. doi: 10.1073/pnas.0703993104.

WHY PARENT TOGETHER?

Marcus William Hunt[1]

[1] Concordia University Chicago, USA

ABSTRACT

The paper offers an account of co-parenthood according to which co-parents are parent and child to one another. The paper begins by reviewing extant theories of the value of being a parent, to see whether the value of co-parenthood is reducible to this. Finding that it is not, I briefly elaborate a theory of parenthood on which parents are those who create persons. Using Aristotle's four causes as a helpful prism, I outline how parents are the cause of their child, and how in causing a child together co-parents become parent and child to one another. For instance, since parents create children by offering themselves as models to be copied, co-parents should enjoy the best type of friendship with one another, each treating the other's flourishing as a human person as their end. I suggest that co-parenthood contains parenthood virtually, that the co-parents' love of their child is a manifestation of their love for one another, that the teleological fulfilled state of the friendship between parent and child exists in the friendship of co-parents.

Keywords: *Aristotle; co-parent; family ethics; parent; solo-parent.*

Introduction

It seems that there is some value, for the parent, in having a co-parent. It seems that this value must be something other than utility of the most basic kind; that financial costs and household chores are split. The value of such things is generic, it could easily be substituted by a financial windfall or hired help. What is the special value of the co-parental relationship? What distinctive contribution to human flourishing does having and being a co-parent make?

One way of answering this question would be to reduce the value of being a co-parent to the value of being a parent. Perhaps having a co-parent is valuable just because it allows one to attain more easily, reliably, or completely, the value involved in parenting—something very valuable. I argue against this type of answer by reviewing the extant theories of why parenting is valuable for the parent. I show how, on these theories, the value of parenting is just as accessible to a solo-parent as a co-parent.

If the value of having a co-parent is not reducible to the value found in being a parent, then close attention to the nature of the co-parenting relationship—scant in the philosophical literature (Cutas and Hohl 2021)—will be needed to discern its value. Since co-parents are those who parent with another, no such account could be entirely agnostic about the nature of parenting. So, I outline my own theory of parenting; parenting is the action that aims to create a person. I then offer an account of co-parenthood on which co-parents are those who stand in relation to one another as both parent and child—that for A to be the co-parent of B is for A to be both the parent and the child of B. I show how, in creating a child together, co-parents take on the goals of creating, and being created by, one another. The co-parental relationship is, in the ideal case, the perfection or completion of the parent-child relationship, the model to which it aspires, in which the child has been brought to maturity and reciprocates the person-creating action of their parent as their now-equal.

The question of the value of the co-parental relationship has some social importance given the growth of solo-parents, those who make the intentional decision to become parents by themselves (as distinct from single-parents, those who find themselves parenting alone due to bereavement, abandonment, etc.). Solo-parents have some media profile (Brockes 2018; Roberts 2019) and online support communities ("Single Mother's by Choice Forum" 2019; "Choice Moms Discussion Boards" 2019). The reports of the United Kingdom's Human Fertilisation and Embryology Authority give some idea of the popularity of solo-parenting. In 2019, 2% of those who undertook *in vitro* fertilization registered as

having no partner, 1,470 people. In the same year, 18% of those who underwent *in utero* insemination with donor sperm registered as having no partner, 1,027 people ("Fertility Treatment 2019: Trends and Figures" 2021). One might also note the report of the United Kingdom's Ministry of Justice that, in 2016, 16.3% of adoption orders were issued to sole applicants, 951 people (MoJ 2019).

The focus of this paper is axiological (value, good) rather than normative (ought, obligation). So, though I do elucidate something valuable that solo-parents necessarily cannot avail of, I offer no ethical prescriptions concerning solo-parenting. My theory employs Aristotelian ethical and metaphysical concepts. The goal of the paper is not to do Aristotle scholarship or reproduce Aristotle's accounts of the family (Aristotle 1991a, Bk. 7; 2011, Bk. 8), but to give a plausible account of the axiology and metaphysics of co-parenthood.

1.   Extant Theories of Why Parenthood Is Valuable for The Parent

I now review the main extant theories of parental rights, each of which involves a theory of why parenthood is valuable for the parent (the philosophical literature focuses much more on the former). I examine whether, on these theories, the value of having a co-parent can be reduced to, explained purely in terms of, some tendency to help achieve the value of being a parent.

1.1   Liberty Theories

For authors such as Charles Fried and William Galston, the rights of a parent are just the rights of a free citizen. Fried writes:

> The right to form one's child's values, one's child's life plan and the right to lavish attention on that child are extensions of the basic right not to be interfered with in doing these things for oneself. (Fried 1978, 152)

Galston writes:

> (…) the ability of parents to raise their children in a manner consistent with their deepest commitments is an essential element of expressive liberty. (Galston 2002, 101–2)

On these theories, being a parent is valuable because it affords an opportunity to live out one's life in ways consonant with one's own values

and commitments without external interference. A solo-parent is just as able to take advantage of this opportunity as a co-parent, and so to attain the relevant value.

Although Fried and Galston have in mind the free citizen's liberty rights as held against the state, their theories suggest that being a solo-parent is more valuable than being a co-parent. Co-parenthood often involves some limitations on the exercise of liberty as compared with solo-parenthood. For instance, one co-parent might insist that the other not eat sugary snacks in front of the child, or not share their interest in blood-sports. These are limitations on liberty that a solo-parent does not face. These limitations are certainly normatively different than those imposed by law. Yet, *ex post* to voluntarily becoming a co-parent, one's freedom is in fact limited in a variety of ways.

1.2   Shaping Theory

Edgar Page argues that the parent's rights are grounded in the value, for the parent, of engaging in a certain action, "shaping":

> (…) parents have a positive desire to influence the course of a child's life, to guide the child from infancy to maturity, a desire to mould it, to shape its life, to fix its basic values and broad attitudes, to lay the foundations of its lifestyle, its priorities, its most general beliefs and convictions, and in general to determine, to whatever degree is reasonable and possible, the kind of person the child will become. (Page 1984, 195–96)

On Page's theory, parenthood is valuable for the parent because it allows them to engage in shaping. A solo-parent is just as able as a co-parent to engage in shaping.

Page's theory implies that solo-parenthood is the more valuable form of parenthood for the parent. At some margins, a solo-parent will be better able to shape the child just as they please. Co-parenthood means neither parent seeing the child shaped in precisely the way that they would prefer. For example, your child spends more time watching TV than you prefer when your co-parent is looking after them, your child is exposed to your co-parent's religious or political attitudes that differ from your own.

1.3   Identity Theory

Yonathan Reshef emphasizes the value of one of shaping's products; a sense of identity between parent and child:

> (…) parents reproduce some of their characteristics in their children and thereby establish a powerful sense of interconnectedness and continuity between their own identity and their child's. Through the intimate process of upbringing parents can bequeath their cultural, national, and religious horizons to their children. Children acquire their parents' language, they are raised according to their parents' values and beliefs, and they follow their parents' practices. Some of the parents' more personal characteristics also pass on to their children, such as favourite dishes, leisure activities, hobbies, body language and outward 'look'. (Reshef 2013, 140)

A solo-parent is just as able to generate a sense of identity between themselves and their child. From Reshef's examples we can see that his view also suggests that solo-parenthood is potentially more valuable for the parent. Co-parents are often from different cultural or religious groups, and will almost always have differing tastes and preferences. At the margins, co-parents will be less able to generate a sense of identity between themselves and their child, or that sense of identity will be weaker than if they passed on only their values, beliefs, nationality, hobbies, and the like.

1.4   Fiduciary Theories

On fiduciary theories, the rights of parents derive from the rights of children. Parental rights are awarded because having these rights helps parents to act as fiduciaries who secure the rights of their children. On this view, parenthood is valuable primarily because it secures the rights of children. Fiduciary theories generally do not involve a specific theory of why parenthood is valuable for the parent, why anyone would want to be such a fiduciary. Jeffrey Blustein's fiduciary theory allows that parents may find many varied kinds of value in being a parent, e.g., achieving a kind of personal immortality, taking pleasure in altruistic behaviour, having a sense of competence, fulfilling an important social role, and the eventual friendship, gratitude, and support, of their adult children (Blustein 1982, 148–50, 175–95). For Blustein, the value of parenthood for the parent is largely subjective. Yet, whatever the value of parenthood might be for the parent, attaining that value depends on them securing the rights of their child. For instance, to attain the valued sense of competence, the parent must ensure that the child is educated appropriately.

Fiduciary theories could explain the value of having a co-parent if, empirically, co-parents are able to secure the rights of their child more reliably or completely than solo-parents. Many studies suggest that the

children of single-parents stand at a higher risk of ills such as psychological problems and poor educational outcomes, even after controlling for financial disadvantage (Lipman et al. 2002; Weitoft et al. 2003). Plausibly, some of these outcomes are violations of the child's rights or are empirically associated with violations. This gives reason to think that solo-parenthood is a less reliable way of attaining the value of parenthood.

Yet, the empirical literature on the children of solo-parents is uniformly positive; such children appear to do as well as the children of co-parents with respect to psychological problems, educational achievement, and the like (Golombok, Tasker, and Murray 1997; Murray and Golombok 2005b; Golombok et al. 2016; Chan, Raboy, and Patterson 1998; MacCallum and Golombok 2004). Plausibly, this undermines reason for scepticism about outcomes for the children of solo-parents. However, this empirical literature features only small-N studies, and at present solo-parenting is done largely by highly-educated and financially stable professionals in their late 30s and early 40s (Golombok et al. 2016). Perhaps solo-parenting is *ceteris paribus* a less reliable fiduciary, or is reliable only in certain socio-economic contexts.

Wherever one stands on these empirical questions, on fiduciary theories the value of having a co-parent would be something highly contingent. Depending on how exactly the value of parenthood is spelled out on a fiduciary theory, one could end up with the result that lacking a co-parent is better than having one. By analogy, a solo-author gets more 'credit' than a co-author, completing a task by myself gives me a greater sense of competence than doing it with another's help, perhaps the parent's sense of personal immortality is stronger if the child is only their own.

1.5   Relationship Theory

Harry Brighouse and Adam Swift offer a theory on which parental rights are partly fiduciary. Additionally, parents have a right to a relationship with their child because it makes "a distinctive and weighty contribution to the well-being of the parent" (Brighouse and Swift 2014, 86). Brighouse and Swift suggest that a successful parent-child relationship depends on some degree of shared values, beliefs, hobbies, and the like (Brighouse and Swift 2014, 155). So, the parent's right to a relationship with the child gives rise to the right to shape the child's values, and the like, to the extent necessary for a successful relationship (Brighouse and Swift 2014, 153).

Conceptually, it does not seem that having a parent-child relationship depends on the child having a relationship of that same type with someone else. Empirically, solo-parents are not found to have worse relationships

with their children. Indeed, one study found "a lower frequency of conflict between mothers and their children in solo mother than in two-parent families" (Golombok et al. 2016, 415).

Plausibly, Brighouse and Swift's theory implies that solo-parenthood is the more valuable arrangement for a parent. If co-parents have different values they will, at some margins, be less able to shape the child's values to the degree necessary for sustaining successful parent-child relationships. On the topic of raising children in a value-neutral way, Brighouse and Swift write:

> The idea that parents should constantly monitor themselves in their relations with their children in order to screen out anything that might have any influence on their children's emerging values is ludicrous. It would risk distancing them, creating artifice in the relationship, and depriving their children of the possibility of the warm, spontaneous, genuine relationship that they need. (Brighouse and Swift 2014, 154)

Yet, to some degree, co-parents do monitor themselves so as not to influence the child's values, and the like, in ways that their co-parent finds objectionable.

1.6   The Work Theory

Joseph Millum offers a theory on which parental rights have two grounds. Some parental rights derive from the role of parents as fiduciaries of children's rights (Millum 2018, 53–56). The other ground is that there are some goods that parenting produces for the parent, that the parent has an interest in enjoying (Millum 2018, 50–53). One is the parent-child relationship. The other is the child's flourishing; their successful development through each stage of childhood, into adulthood. On Millum's theory what bridges from these things of value to parental rights is an investment principle; "the extent of an agent's stake in an entity is proportional to the amount of appropriate work he or she has put into that entity" (Millum 2018, 25). As such, Millum's theory also seems to imply that solo-parenthood is the more valuable arrangement *qua* the value of parenthood—no one else will have a stake in these goods; all the work, all the goods realized, will be one's own.

1.7   Discussion

Each of these theories of the value of parenting is unable to explain why having a co-parent is valuable for the parent. Instead, they suggest some tension between having a co-parent and attaining the value of parenthood. So, by induction, plausibly the value of the co-parental relationship is not reducible to the value of being a parent—the former must have some value of its own.

To bolster this conclusion, note that parenting is clearly an action that can be done by one parent. By contrast, actions such as "warring with Venice", lecturing to a class, or playing in a band, require the participation of many agents to be tokened. A solo-parent performs the same action as the co-parent, and so can attain the same value, the goal that the action aims at.[1] So, any additional value of parenting that co-parenting would tend to promote would be highly contingent (in terms of personalities and socio-economic situations). Such an account would fail to justify or explain the popularity and normative status of co-parenting.

2.   The Person-Creating Account of Parenthood

My view is that parenting is the action that aims to create a person. A parent is one who does this action. A child is the object of this action. This view also allows that the value of parenting is available to the solo-parent, but suggests an account of the metaphysics and value of co-parenthood. I now outline this action.

Actions are defined and differentiated by their goals (Aristotle 2011, 1094a; Wilson and Shpall 2012). To do a particular action is to will a particular goal, to use *energeia* for its *ergon*, to be active in its operation (Aristotle 2011, 1098a 7-18, 1106a 23). Muttering to yourself and talking to someone else may involve identical behaviours, yet are different actions due to their different goals. Although parents and non-parents often behave in the same ways, their actions are different. For example, a parent and a school canteen-worker both give the child food. Parents engage in this behaviour because it contributes to their person-creating goal. Lacking this

---

[1] One objection here might be to deny that solo-parents and co-parents perform the same action. Perhaps the action of a co-parent is a collective action of the co-parents. In response, I would deny that co-parents do a different action than the solo-parent. Such a thesis does not explain the normative contours of parenthood—e.g., that each parent is able to exercise the normative powers of parenthood as an individual, that the ethical duties of a co-parent to their child are not different from those of a solo-parent (Hunt 2022).

goal, having some other goal, the school canteen-worker is not doing the action of a parent.

Since the goal of the parent is to create a person, the distinction between "biological parenthood" and "social parenthood," "bearing and rearing" (Archard 2004, 137) can distract from the underlying unity of the parental action. There are different ways of participating in person-creating action. One parent participates in it by bringing a new human organism into existence—contributing gametes, gestating—and another by perfecting that organism's existence—feeding them, educating them. Here, nutrition and child-rearing continue the action that procreation initiated. Though the behaviours of each part of the parental action may be very different, they are parts of the same action since they have the same goal. Since solo-parents can be both biological and social parents, they can do every part of this action.

Action aims at the good (Aristotle 2011, 1094a 1). To act is to approach the good under the guise of a goal; willing to make the world better, creating something of value—whether a good artefact or a good state of affairs, internal or external to the agent. An action, and the creation that it aims to actualize, can be more or less perfect in two ways; *fulfilled* and *finished*.

An action or creation that is more fulfilled is one of the better tokens of its type, exhibiting more of its characteristic value. For example: one speech is rousing and another not, one painting is more beautiful than another. For a creator to fulfil a creation, to make it better, is for them to continue in the act of creating it, rather than a distinct action. The notion of fulfilment rests on the Aristotelian and Platonic doctrine of the interconvertibility of being and well-being, existence and goodness (Aristotle 1991b, 1051a 18-21; Plato 1997c, 508b-e). To flourish is to exist more, whilst to wither—to be diseased, ethically vicious, ignorant, friendless—is to tend toward non-existence, showing forth human form less fully. So, we rightly use intensives for the better painting (action and artefact): "now *that* is (a) painting!"

As things that become over time, an action and its creation can be more or less finished. Consider *Sagrada Familia*. This is an unfinished cathedral, rather than a non-cathedral. To finish something is a part of creating it. Now that *Sagrada Familia* exists in the minimal sense that the sufficient condition for being that cathedral is met, the continued sculpting, hoisting, designing, etc., of the workers seamlessly continues the act of its creation.

Parents seek to perfect their child in both of these senses. Parents, in fulfilling and finishing an already-existing human person, continue their participation in the parental action. The person-creating action does not end at conception, or birth, or for many years after. As a person is a substance of a rational nature (Teichman 1985), a good human person is one who makes their animality participate in their rationality. A good human person enjoys some knowledge of the world around them, basic skills like speech, balanced affect, prudence, a healthy body, a family and community, and, most especially, ethical virtue.

I now characterize the parenting action in terms of Aristotle's four causes, four respects in which one person can create another (not that this is the only conceptual schema that could be fruitfully applied to it). I will then be able to show how co-parents are parent and child to one another in these four ways. Throughout, I speak of what parents and children do in the ideal case, rather than making empirical claims.

2.1   Parents as Material Causes

Matter is potential, the potential for receiving form (Aristotle 1991b, 1029a 20-26, 1036a 7, 1048b 1-7); matter is "a receptacle of all becoming—its wetnurse" (Plato 1997d, 49a), itself lacking form, but able to manifest it. Form is pattern, activity, and goal (respectively Aristotle 1991b, 1013a 26, 1050a 15-22, 1023a 32; Plato 1997a, 389b; 1997c, 508d-e; 1997b, 75a), matter is that in and from which form appears.[2]

In generation, in creating after their kind, children are made from the matter of their parent's bodies, their gametes. In generation, parents are matter in relation to their child in that they are, together, the potentiality out of which a new child can emerge, they are that which can receive the actuality of new life.

In nutrition, parents continue as material causes of their child by providing them with more matter to incorporate—at first from the mother's body via the umbilical cord, then via lactation, and then, proximately, through other food. Nutrition helps actualize the child's potential for growth. In child-rearing, parents are material causes of their child in that they provide the

---

[2] "Form" and "formal cause" are not quite interchangeable. For both Aristotle and Plato, "form" is the formal, efficient, and final cause of matter (pattern, activity, goal). Aristotle's discussion of the four causes conceptually separates out these three aspects of form. This Aristotelian-Platonic conception of matter differs from the more modern conception of matter as that which is fundamentally real, the fundamental particles. The former prefers the metaphor of form descending into matter, the latter the metaphor of form emerging out of matter. These understandings overlap in seeing matter as an indefinite sludge which, by being properly arranged, becomes otters, cassette tapes, and everything else.

potentials out of which their child's perfection is actualized—e.g., in buying a xylophone, parents help actualize their child's potential musical skill. Of particular importance is that the parent's own life (as opposed to various objects and environments that they might orchestrate) is a material cause of the child. A parent's body and mind are actualized in some ways and not others by their parenting action—changing diapers, learning to tolerate messiness, practicing patience, etc. These states of the parent are the parent as enformed by their child, as matter actualized by the child as form, the child's life received into their own. The parent's own life is, then, a potential in which and from which the child is created and perfected.

Lastly, parents are material causes of their child in the sense that they attend to their child considered as a material thing, as a body, that has the potential for change, including the potential for non-existence—keeping them warm, washing them, clipping their nails.

As actions are defined by their goals, not everyone or everything who happens to be a material cause of a child's existence and perfection counts as their parent. Gametes are not parents, the formula salesman is not a parent, and the person who generously donates formula is not a parent, since the potential that they aim to actualize is more limited in scope than that of the parent (they are not offering to rock the baby to sleep, take them to visit grandma, educate them, etc.). This point applies to each of the four causes.

2.2   Parents as Formal Causes

The formal cause is the pattern according to which a goal actualizes matter (Aristotle 1991b, 1013a 26). In generation, parents impart the human form to their child. The parent's biological contribution determines the organism as one that grows according to the human pattern, that in time will more fully manifest the human pattern.

In child-rearing, parents act as formal causes of their child by using themselves as the model, the pattern, from which their child learns the capacities of human personhood, such as speech, social skills, and decision-making—playing with them, exhibiting appropriate affect, showing them how to make eggs. This is not to say that parents propose themselves to their child as perfect models of human personhood, as ideals to be copied in every respect. Rather, parents use themselves as the actual models of human personhood for their child to copy, and with an awareness that their child will often copy and idealize them (which imposes a normative demand to model sufficiently well).

Here, the difference of my theory of parenthood from Page's is best captured in his statement that the goal of the parent is to determine "the kind of person the child will become" (Page 1984, 196)—a lover of crochet, a trade-unionist, a fan of *The Smiths*. On my account, the goal of the parent is to create a human person, with no 'kind' being selected for other than 'good'. Shaping a child toward various particularities (usually, those of the parent) is often harmless or good, but not the goal of a parent as such. Again, in generation parents do not just impart the generic human form to their child, but their own individual human form—a set of physical and psychological particularities. My account claims that being a biological parent is valuable because it means participating in a part of the person-creating action, but not because of these particularities.

Lastly, parents act as formal causes of their child in the sense of attending to them *qua* formal cause, as a being that should be actualized according to a certain pattern, that can deviate from it—attending to their health, ethical behaviour, psychological state.

2.3   Parents as Efficient Causes

Efficient causes bring about change (Aristotle 1991b, 1013a 24), they manifest a new activity in time. In generation, parents cause a change; a human person comes to be. In nutrition and child-rearing, parents continue as efficient causes of their child by acting in ways that keep them in existence and bring about their increasing perfection as a human person.

2.4   Parents as Final Causes

A final cause (goal, end, telos) is "that for the sake of which a thing is" (Aristotle 1991b, 1013b 3), that toward which it is ordered, that in light of which matter is actualized according to a given pattern. To explain how it is that the parent is *the* final cause of the child, I must first elaborate on how it is that the child is *the* final cause of the parent. The parent makes the child *a* final cause—wills their flourishing for its own sake, dedicates some portion of their energeia to this ergon, just as they might do for a friend, dog, or political cause. More than this, the child's existence and perfection as a human person is the goal of a parent in that this goal is sufficient to justify the parent's continued existence and growing perfection. A parent is willing to give up their weekends for their child's sake, to quit smoking, work a dead-end job, to die, or continue living, for their child's sake—to *be* for their sake.

Whilst radically altruistic, the parent is simultaneously radically selfish–the child is the parent's final cause because the parent treats their own existence and perfection as their final cause. From a biological perspective, considered as an organism, the parent flourishes by having a child;

> The most natural act is the production of another like itself, an animal producing an animal, a plant a plant, in order that, as far as its nature allows, it may partake in the eternal and divine. (Aristotle 1991c, 415a 26-30)

Having a child is not merely *a* goal of the organism, but *the* goal of the organism *qua* organism—the goal which distinguishes living things from non-living things is generation, reproduction. In procreating and rearing a child, the organism affirms its own goodness in the manner of an organism.

Since a rational being affirms itself rationally, having a child is also the goal of the parent considered as a person. Our cognitive and conative powers aim at the good. So, *ceteris paribus*, the more valuable the goal, the better it is for the agent that they pursue it. For instance, knowing how many blades of grass there are in a field is not a very valuable goal (except extrinsically, in some bizarre scenario). So, counting the blades of grass is a worthless activity, human languishing. The goal of creating and perfecting a human person is an extremely valuable goal. Accordingly, to create and perfect a human person is the greatest, most transformative, and most sustained, actualization of their potential that most people experience. Some indication of this is that "94% of parents say that having children is worth it despite the costs, and parents report that having children is *the most positive event in their lives*" (Nelson, Kushlev, and Lyubomirsky 2014, 8 *my italics*). Again, whilst Abraham Maslow's classic hierarchy of needs featured "self-actualization" as the peak need, recent evolutionary psychology finesses this to "parenting" (Kenrick et al. 2011). For the human person to have a child is for them to affirm their own goodness in the manner appropriate to a person; by choice, lovingly, and in relationship, casting their self into an other self. As the parent performs their action by being a model of human personhood, it actualizes them toward personhood.

I now begin turning to how, in light of this, the parent is the end of the child. All good things call for a return, for us to respond to them in some way; the friendly dog ought to be pet, the sweet crisp apple ought to be savoured, the chance ought to be seized. One helpful distinction to draw among these responses is that between passive and active responses.

To be passive toward to a good thing is to allow it to be good for you, to be affected by it in the way that is best. For instance; laughing at a humorous joke, using a financial windfall prudently, remembering the interesting fact. To be passive to a good includes cognitively and conatively appreciating its goodness, and also absorbing it, bringing it into one's life. To have an active response to a good is to act to promote the good, to do something for it, typically in a way that is not 'selfish'. For example, preserving the good thing in existence, copying it or creating things that are like it, sharing it with others, honouring it. For example; telling the humorous joke or interesting fact to your friend, pointing out the beautiful sunset to your friend, voting to keep the forest preserve, imitating the display of skill, returning the kind favour.

In cases where the good was produced by a moral agent, a return in the more paradigmatic sense, doing something that is good for that very agent, may be an appropriate active response. However, the agent who does good automatically receives a good return, insofar as "it is a greater perfection for a thing to be good in itself and also the cause of goodness in others" (Aquinas 1947, ST I Q103, A6, co.) and "it belongs to the essence of goodness to communicate itself to others" (Aquinas 1947, ST III, Q1, A1, co.). These dicta also illustrate that passive responses to a good lead to active responses; e.g., that part of what it is to truly appreciate the melody is to be disposed to invite others to hear it.

Each ethical virtue is a disposition to action and passion (emotion, desire) that responds to a particular type of good in the way that is best (Aristotle 2011, 1104b 15). For example, moderation responds to the good of bodily pleasure. The ethically virtuous person displays *ordo amoris* (the order of love); they are disposed to respond to all goods in the way that is best in a given circumstance; they are always poised to achieve what is best (Aristotle 2011, 1106b 5–30) (rather, than, say, responding to bodily pleasure at the expense of health).

Filial piety is the ethical virtue that makes a response to the goodness of the parental action, is the ethical virtue of a child *qua* child. Its passive response, most obvious in the life of an immature child, is to imitate the model of human personhood presented by their parent, to obey they instructions, and to "attach" to their parent; to feel loved by them, to feel safe with them. In these ways, a child is effected in the way that is best by their parent's action. The active response of filial piety, most obvious in the life of a mature child, is to imitate their parent by themselves having children, reciprocating their parent's person-creating action insofar as this is possible (feeding their elderly parent, keeping them socialized, etc.) and, to perfect their parent's action by being a good person (Hunt 2023). In

these ways, the parent-child relationship is one in which each party treats the other as an end by treating themselves as an end, perfects the other by perfecting themselves. Since beings are defined by the goals that they are for the sake of (e.g., the knife is for cutting), and since the child is for the sake of the parent, the existence and perfection of the child is defined with reference to the parent; a token person could not have had different token parents, and piety is "leader of all the virtues" (Hierocles 2002, 174).

The response of filial piety, as an imitation, involves a reciprocation of the parental action—being a material, efficient, formal, and final cause of one's own parents. In other words, whilst the parent-child relationship begins as one in which these two roles are played exclusively by two distinct individuals, in its mature form, in its teleologically fulfilled state, the two individuals are both parent and child to one another. I now illustrate how this is the relationship that obtains between co-parents.

3.   Co-Parents as Parent and Child to One Another

Co-parents are creators of one another and created by one another, and so are both parent and child to one another.

3.1   Co-Parenthood and Final Causes

The goal of the parent is to create and perfect a human person. Parents do this by offering themselves as goals, as models, to their children. Good parents provide a model of human personhood that is worthy of imitation; living well, ethically, beautifully. If a child ends up living well *despite* their parents, then those parents have not parented well.

The co-parent is one who has the same goal in respect to the same child, who also offers themselves as a model to that child. So, whether one co-parent does their action well is of intrinsic concern to the other—if I am showing my child a good way to live, but allow my co-parent to show them a bad way to live, then I am failing in my capacity as a parent: I am not in fact showing them a good way to live. For example, if your co-parent overshares their anxieties with your child, or drinks too frequently, or otherwise sets a bad example, this is intrinsically bad for you, *qua* the model that you propose to your child in allowing these things. So, co-parents have a concern for one another *qua* the child's model, final cause.

To pursue this concern, co-parents treat one another as final causes. That is, they offer themselves to each other as models, receiving one another as models, particularly in relation to child-rearing activities—e.g., "when I

15

get upset by the baby's screaming, I count to 10", "when I'm getting her ready, I let her put her jacket on by herself".

That co-parents treat one another as models makes sense of the claim that co-parents have a kind of say over one another; that they can impose claims of significant normative weight on one another in making requests of one another, that they should deliberate and negotiate together in forging a shared way of life. This seems analogous to the say that parents and adult-children have over one another, who may make and take requests with patience, compromise, and trust in one another's good will, but not issue "commands".

The best modelling of human personhood will involve modelling the best type of human friendship—of mutual love, stability, equality, openness, accountability, ethical elevation, and the like. The goal of the parent is not to create an isolated rational agent, but one who is embedded in, and will embed themselves in, rich personal relationships. The friendship that co-parents enjoy with one another should be of this best type, since it is this model that they propose to their child as being the best. So, in having their child as their final cause, co-parent's have one another as final causes; pursuing one another's human flourishing, treating one another as a sufficient reason for being. At the biological level, the two co-parents again treat one another as final causes in procreating, affirming one another's goodness in the manner appropriate to sexual organisms.[3]

Here we have seen one respect in which co-parenthood originates, in the order of time, in the parental action—we become co-parents by parenting a child together. Yet, in the order of being, the parental action originates in the co-parental action, the former finds its perfection in the latter. The co-operation, mutual love, and mutual say of the co-parents enables and shows itself in the parental action, and calls the child to develop toward co-operation, mutual love, and shared life, with their parents and others. The action of the parent exists virtually in the action of the co-parent (Aquinas 1947, ST I Q4 A2); the former is contained pre-eminently in the latter, the former unveils something borne in the latter. Or again, the co-parental action is the paradigm and the parental action its image, the love toward the child the matter in which the love between co-parents is manifested:

---

[3] Within current technology——i.e., excluding cloning——generation is something that requires two parents; the two parents are together the final, formal, efficient, and material cause of the child. I think it is best to describe gamete donors, surrogate gestators, or those who place their child for adoption, as ceasing to be parents because their participation in the parental action ceases. At some point in time, their goal and action was that of a parent (even if this was never their conscious intention). So, there are no solo-parents in the sense "the only parent that a child has ever had", but there are solo-parents in the sense "by choice, the only parent that a child has and will have".

"the giver possesses primitively the character which it gives, while the recipient is by derivation what the giver is" (Proclus 1963, Pr. 18).

3.2   Co-Parenthood and Formal Causes

Co-parents recognize one another as formal causes of the same child. So, each co-parent enables the other to act as a formal cause—to impart their preferences, tastes, and dispositions with the child. Yet, since the action of the parent is to perfect their child's formal cause—to shape them in the way that is best, towards the full expression of human form—co-parents are also formal causes of one another. That is, co-parents offer and receive one another's form; harmonizing their way of life, their evaluative attitudes, and so forth.

For example, if my formal cause has departed from the human—e.g., if I am viciously jealous—then, as a parent, my co-parent has an interest in repairing this, imparting the human form upon me by their example, encouragement, and correction, and as a parent I have an interest in being receptive to their action. Again, given the scope of the parental action, that a parent is a model of human personhood (not just a model of cookery, etc.), co-parents have an interest in one another as models of human personhood. So, in respect of formal causes, co-parents are related as creators and creations of one another's human personhood, as parent and child.

3.3   Co-Parenthood and Efficient Causes

In child-rearing, co-parents, recognizing one another as parents, enable one another to be efficient causes of their child's perfection; ensuring that the child learns interests from both, hears the perspectives of both, etc. For example, the "primary caregiver" might arrange the child's daily schedule in a way that maximizes the time that the child spends with the "secondary caregiver", "working-parent", or one co-parent might go along on a fishing expedition if this enables the other to share their interest in fishing with the child. As the action of the parent touches on practically every aspect of their life, and can reasonably be expected to last until their death, co-parents act as efficient causes of their co-parent's perfection *qua* human person in acting as efficient causes of their child's perfection. In harmonizing their way of life to the degree necessary for performing the parental action well, in treating one another as final and formal causes, co-parents change, and are changed by, one another. As children are receptive to being changed by their parents, so should co-parents be. Negatively, imagine that one had a very bad co-parent—in light of the burdens of parenting, this would seriously impede one's human flourishing, make it

hard to change one's life for the better. So, as efficient causes of one another's perfection as human persons, co-parents are related as parent and child.

3.4   Co-Parenthood and Material Causes

As noted, parents are the material cause of their child in that they are matter for the child's form, potentials in and from which the child is actualized. The rhythms of the child's way of life are incorporated into that of the parent—the parent wakes at one time rather than another, plays games rather than working. The respective patterns of life of co-parents are together a material cause of the child—e.g., knowing that my co-parent will attend to dinner, I can do the laundry. For this reason, co-parents are material causes to one another as human persons. The pattern of life of each co-parent is expressed in, and co-ordinated with, the other. By harmonizing their patterns of life, each provides potential for the other to flourish as a human person. Negatively, imagine the worst co-parent—co-parenting with such a one would very seriously limit your potential for flourishing. Again, co-parents are material causes of one another in the sense of attending to one another *qua* material thing—caring for one another's nourishment, health, etc. So, co-parents are material causes to one another as human persons, and so are related as parent and child.

3.5   The Value of Having a Co-Parent

With this description of co-parenthood in hand, its value is self-evident. Co-parenthood is a form of friendship, and friendship in general is valuable for creatures like us. The friendship of parent for child, and of child for parent, are generally acclaimed as the greatest human friendships, the greatest bestowals and receipts of goodness. Co-parents, as parent and child to one another, share a friendship that is the ontological ground and teleological completion of these friendships. Co-parenthood is a friendship which unites human persons as human persons through the act of creating a human person. It is an extensive friendship, touching every aspect of the friend, and an intensive friendship, touching them at the deepest level, a friendship that unites and transcends its participants, a friendship whose fruit is an image of the divinity, a human person.

To be sure, it seems that two people could simply decide to act toward one another in the way that co-parents should—using one another as models, caring for one another, etc. However, the co-parental action is distinguished from such a friendship in that it has an objective basis in the child; given the responsibilities of each co-parent toward the child, their

responsibilities toward one another are not a matter of ongoing voluntary decision. I now turn to answering objections.

4.   Responses to Objections

(i)   To say that co-parents are parents to one another and children to one another seems to conflict with the ordinary use of these terms—if asked to point out their parents or their children, no one would point out their co-parent.

I hope that the claim that co-parents are parent and child to one another is somewhat surprising. A philosophical account should produce insights that are not commonplace. It makes sense that in pointing out "child" and "parent" we would point to uncomplicated cases. So, the present objection is not troubling. My account does indeed revise the common-usage extension of "child" and "parent" —my application of these terms to co-parents is in no way metaphorical or analogical. But, I do not think that my account revises the common-usage intension of these terms (that parents are creators, children the created).

(ii)   When you say "co-parents" don't you really mean "marital partners"? The features that you ascribe to the co-parental relationship seem very similar to marriage.

In common usage, "co-parents" is often used to pick out those who are only or primarily co-parents; divorced people, or platonic parents. Yet, most people who are co-parents are also romantic partners or marital partners, and this seems to be the normative ideal for most people. It might be that the marital relationship involves, is ordered towards, having children, and therefore the co-parental relationship. However, it seems unlikely that the marital relationship (which I have not yet considered in philosophical detail) and the co-parental relationship are identical. The extensions of the two can be diverge, even if one takes the view that their extensions ought, ideally, to coincide. *Prima facie*, it seems plausible to me that the marital relationship is architectonic in relation to the co-parental relationship—includes it but supersedes it; that there are aspects of the relationship between marital partners that are not exhausted in the co-parental relationship, that co-parents as such do not *a fortiori* count as marital partners.

(iii)   If co-parenthood is so valuable, what explains the rise of solo-parenthood?

A survey of users (n=290) of *Choice Moms,* an online community for solo-mothers, was conducted to investigate their reasons for wanting to have a child. Naturally enough 97% cited "I wanted to be a mother", 62% cited "I was getting older", 50% cited "I was financially secure", and 43% cited "It was now or never" (Jadva et al. 2009, 179). 76% of the users reported having been in a long-term relationship in the past. As their reason for not having had a child during their previous long-term relationship, 64% cited "The relationship was not right", 48% cited "The timing was not right" and 26% cited "Partner did not want a child" (Jadva et al. 2009, 177–78). In a smaller survey (n=27) of heterosexual solo-mothers, 87% reported that they would like to become romantically involved with a man in the future, with 40% citing the desire for the child to have a father as a reason for having a romantic relationship in the future (Murray and Golombok 2005a, 250).

Whilst limited and ambiguous, this evidence suggests that many solo-parents might agree that co-parenthood is valuable. It seems that the choice for most solo-parents is between solo-parenthood and no parenthood. As seems reasonable, they did not make the perfect the enemy of the good. The demographics of solo-parenting suggest that sociological and economic factors are the key drivers of the rise in solo-parenting, not a perception that it is axiologically interchangeable with, or superior to, co-parenting.

(iv)   Don't solo-parents also have to co-create their child with others? Single-parents often engage their child with extended support networks, including the parent's own parents, siblings, close friends, and neighbours, as well as the child's daycare providers, teachers, coaches, and other kinds of mentors.

In response, I deny that solo-parents co-create their child with these others. As creating a child just is the action that makes one a parent, to say that these others were co-creating their niece or student would be to say that they were their parents. This is counter-intuitive. Whilst extended support networks add many good things to a child's life—an interest in the arts, an appreciation for some personal virtue—the goals of the other agents in these cases are more limited than the goal of a parent—to help them in some respect, to create some kind of trait in them, not to create *them*. This is not to deny that, in many cases, a member of the child's extended support network may become their parent—e.g., the grandparent who replaces an

absent parent—but then we have a case of co-parenthood. I note that none of these types of relationship are closed off to co-parents or their children.

(v)   Doesn't this view suggest that the more co-parents the better? Why not 3 or 4 co-parents?

To state my intuitions, I am comfortable with the claim that more than 2 people can be co-parents, and find the value thereof, but uncomfortable with the claim that "multi-parenthood" is better than "duo-parenthood".

The value of having a co-parent is that in this type of friendship one is charged with helping to create, and simultaneously be created by, another human person. As we noted in the first section, the person-creating action is one that can be done by an individual: co-parents and solo-parents do the same action to their respective children, and are both able to attain the relevant value. So, *qua* the value of parenting, there is no advantage to the co-parent in doing their parenting action with 3 or 4 others, rather than just with 1 other, or by themselves. To deny this would be to deny that solo-parents are in an axiologically equal position to co-parents vis-à-vis parenting. Likewise, *qua* the valued of being parented, to conclude that multi-parenting is better for the child than duo-parenting one would also have to conclude that duo-parenting is better than solo-parenting (something that, I suspect, liberal-minded advocates of multi-parenting might not wish to conclude).

A separate response can be drawn from Richard Swinburne's *a priori* argument for the Trinity (Swinburne 2018, 430). Although love can obtain between a dyad, a love that is not jealous is more perfect, a love that shares the beloved with others is more perfect. God is the perfect being. So, God must be a triad rather than a dyad. Yet, God need not be more than 3 persons since more than 3 would not make the love between the persons less jealous or more sharing: 4 would be otiose. Translated to our context, the child shares the role of the Holy Spirit, the giver of life; they allow the love between the two co-parents to not be jealous, but to be shared with another. So, to introduce more co-parents would be otiose, would not make co-parenthood more perfect.

5.   Conclusion

To give an account of the value of having a co-parent I first examined whether such value could be reduced to the value of being a parent, and found that it probably cannot. So, I gave an account of parenthood and co-parenthood. The view that emerged is that parents are those who create

persons, and that co-parents, in creating a person together, come to create and be created by one another, are parent and child to one another. I noted that the idea that the parent-child relationship is a reflection of the co-parental relationship meshes with the traditional Aristotelian and Platonic view that effects pre-exist in their causes, are contained virtually in their causes (Aquinas 1947, ST I Q4 A2 co., Q19 A4 co.).

Acknowledgments

REFERENCES

Aquinas, Thomas. 1947. *Summa Theologica*. Translated by English Province Dominicans. New York: Bezinger Bros.

Archard, David. 2004. *Children: Rights and Childhood*. Second Ed. Abingdon: Taylor & Francis.

Aristotle. 1991a. "Eudemian Ethics." In *The Complete Works of Aristotle, Volume 2*, edited by Jonathan Barnes. Princeton, NJ: Princeton University Press.

———. 1991b. "Metaphysics." In *The Complete Works of Aristotle, Volume 2*, edited by Jonathan Barnes. Princeton, NJ: Princeton University Press.

———. 1991c. "On the Soul." In *The Complete Works of Aristotle, Volume 1*, edited by Jonathan Barnes. Princeton, NJ: Princeton University Press.

———. 2011. *Nicomachean Ethics*. Translated by Robert Bartlett and Susan Collins. Chicago: University of Chicago Press.

Blustein, Jeffrey. 1982. *Parents and Children: The Ethics of the Family*. Oxford: Oxford University Press.

Brighouse, Harry, and Adam Swift. 2014. *Family Values: The Ethics of Parent-Child Relationships*. Princeton, NJ: Princeton University Press.

Brockes, Emma. 2018. "Going It Alone: Why I Chose Single Motherhood." The Guardian. https://www.theguardian.com/lifeandstyle/2018/jun/23/going-it-alone-why-chose-single-motherhood.

Chan, Raymond W., Barbara Raboy, and Charlotte J. Patterson. 1998.

"Psychosocial Adjustment among Children Conceived via Donor Insemination by Lesbian and Heterosexual Mothers." *Child Development* 69 (2): 443–57.

"Choice Moms Discussion Boards." 2019. Choice Moms. 2019. http://www.choicemoms.org/discussion-boards/.

Cutas, Daniela, and Sabine Hohl. 2021. "In It Together? An Exploration of the Moral Duties of Co-Parents." *Journal of Applied Philosophy* 38 (5): 809–23.

"Fertility Treatment 2019: Trends and Figures." 2021. London. https://www.hfea.gov.uk/about-us/publications/research-and-data/fertility-treatment-2019-trends-and-figures/.

Fried, Charles. 1978. *Right and Wrong*. Cambridge, MA: Harvard University Press.

Galston, William A. 2002. *Liberal Pluralism: The Implications of Value Pluralism for Political Theory and Practice*. Cambridge: Cambridge University Press.

Golombok, Susan, Fiona Tasker, and Clare Murray. 1997. "Children Raised in Fatherless Families from Infancy : Family Relationships and the Socioemotional Development of Children of Lesbian and Single Heterosexual Mothers." *Journal of Child Psychology and Psychiatry* 38 (7): 783–91.

Golombok, Susan, Sophie Zadeh, Susan Imrie, Venessa Smith, and Tabitha Freeman. 2016. "Single Mothers by Choice: Mother – Child Relationships and Children's Psychological Adjustment." *Journal of Family Psychology* 30 (4): 409–18.

Hierocles, of Alexandria. 2002. "The Commentary of Hierocles the Philosopher on the Pythagorean Verses." In *Hierocles of Alexandria*, edited by Hermann S. Schibli. Oxford: Oxford University Press.

Hunt, Marcus William. 2022. "Parental Compromise." *Critical Review of International Social and Political Philosophy* 25 (2): 260–80.

———. 2023. "Person-Creating and Filial Piety." *The Journal of Value Inquiry*. Forthcoming.

Jadva, V, S Badger, M Morrissette, and S Golombok. 2009. "'Mom by Choice, Single by Life's Circumstance . . .' Findings from a Large Scale Survey of the Experiences of Single Mothers by Choice." *Human Fertility* 12 (4): 175–84.

Kenrick, Douglas, Vladas Griskevicius, Steven L Neuberg, and Mark Schaller. 2011. "Renovating the Pyramid of Built Upon Ancient Foundations." *Perspectives on Psychologlical Science* 5 (3): 292–314.

Lipman, Ellen L, Michael H Boyle, Martin D Dooley, and David R Offord. 2002. "Child Well-Being in Single-Mother Families." *Journal of the American Academy of Child and Adolescent Psychiatry* 41 (1): 75–82.

MacCallum, Fiona, and Susan Golombok. 2004. "Children Raised in Fatherless Families From Infancy : A Follow-Up of Children of Lesbian and Single Heterosexual Mothers at Early Adolescence." *Journal of Child Psychology and Psychiatry* 8: 1407–19.

Millum, Joseph. 2018. *The Moral Foundations of Parenthood*. Oxford: Oxford University Press.

MoJ, Ministry of Justice. 2019. "Family Court Statistics Quarterly: Data Visualisation Tool." London. https://public.tableau.com/profile/moj.analysis#!/vizhome/FamilyCourtstatisticstool2019Q2/Frontpage.

Murray, Clare, and Susan Golombok. 2005a. "Going It Alone: Solo Mothers and Their Infants Conceived by Donor Insemination." *American Journal of Orthopsychiatry* 75 (2): 242–53.

———. 2005b. "Solo Mothers and Their Donor Insemination Infants: Follow-Up at Age 2 Years." *Human Reproduction* 20 (6): 1655–60.

Nelson, S. Katherine, Kostadin Kushlev, and Sonja Lyubomirsky. 2014. "The Pains and Pleasures of Parenting: When, Why, and How Is Parenthood Associated with More or Less Well-Being?" *Psychological Bulletin* 140 (3): 846–95.

Page, Edgar. 1984. "Parental Rights." *Journal of Applied Philosophy* 1 (2): 187–203.

Plato. 1997a. "Cratylus." In *Plato: Complete Works*, edited by John M. Cooper and D. S. Hutchinson, 101–56. Indianapolis: Hackett.

———. 1997b. "Phaedo." In *Plato: Complete Works*, edited by John M. Cooper and D. S. Hutchinson, 49–100. Indianapolis: Hackett.

———. 1997c. "Republic." In *Plato: Complete Works*, edited by John M. Cooper and D. S. Hutchinson, 971–1223. Indianapolis: Hackett.

———. 1997d. "Timaeus." In *Plato: Complete Works*, edited by John M. Cooper, 1224–91. Indianapolis: Hackett.

Proclus. 1963. *The Elements of Theology*. Edited by E. R. Dodds. Oxford: Oxford University Press.

Reshef, Yonathan. 2013. "Rethinking the Value of Families." *Critical Review of International Social and Political Philosophy* 16 (1): 130–50.

Roberts, Genevieve. 2019. *Going Solo: My Choice to Become a Single Mother Using a Donor*. London: Piatkus.

"Single Mother's by Choice Forum." 2019. Single Mothers by Choice. 2019. https://forums.singlemothersbychoice.org/.

Swinburne, Richard. 2018. "The Social Theory of the Trinity." *Religious Studies* 54 (3): 419–437.

Teichman, Jenny. 1985. "The Definition of Person." *Philosophy* 60 (232): 175–85.

Weitoft, Gunilla Ringbäck, Anders Hjern, Bengt Haglund, and Måns Rosén. 2003. "Mortality, Severe Morbidity, and Injury in Children

Living with Single Parents in Sweden: A Population-Based Study."
The Lancet 361: 289–95.

Wilson, George, and Samuel Shpall. 2012. "Action." In *Stanford Encyclopedia of Philosophy*. Stanford University Press.

MORALITY WITHOUT CATEGORICITY

Elizabeth Ventham [1]

[1] University of Salzburg, Austria

ABSTRACT

This paper argues that an agent's moral obligations are necessarily connected to her desires. In doing so I will demonstrate that such a view is less revisionary—and more in line with our common-sense views on morality—than philosophers have previously taken it to be. You can hold a desire-based view of moral normativity, I argue, without being (e.g.) a moral relativist or error theorist about morality. I'll make this argument by showing how two important features of an objective morality are compatible with such a desire-based account: 1) morality's authoritative nature, 2) our ability to condemn immoral agents.

Keywords: *meta-ethics; practical reasoning; hypothetical imperatives; desire; moral realism.*

Introduction

It's common to think of some of our normative reasons and obligations as being *moral*. Bren, for example, might have a moral reason to attend a protest in her city. We can suppose it's an important cause, and one where anyone expressing their support will be doing something positive. We can also suppose that Bren is a good person, who cares about the cause. But what exactly gives Bren this reason? And if we subscribe to a theory that explains Bren's (practical, normative) reasons as being necessarily contingent on her desires, does that mean that her reasons can't really be *moral* reasons?

This paper argues that a desire-based theory of normativity need not lead us away from moral realism. Furthermore, it argues that even though this means people without the right moral desires won't have moral obligations, this account is still less revisionary than it might seem. I begin by explaining the positive view (that of desire-based normativity) in section 1. In section 2 I will introduce two kinds of opponent: advocates for desire-based theories of reasons, who take this to be evidence that moral realism is in some sense a mistake, and those who are committed instead to the truth of morality, and who think that the desire-based view, therefore, must be wrong.

In section 3 I will make my argument against this false dichotomy by addressing two different features that seem to be important for morality to have, and that my opponents might believe are incompatible with a desire-based view of normativity. The first feature is morality's inescapable and authoritative nature. After all, a key component of morality seems to be that it is in some sense objective and that it holds some power over us. I respond here in two ways: firstly, just because the normativity of morality is conditional on agents having certain desires, that doesn't mean that we can escape the hold that morality has over us any more than we can 'escape' our own desires. I cannot stop myself wanting to be good, or wanting to make other people happy, for example, just because it would be convenient. Secondly, I make an important distinction about what it means to say that normativity is desire-based. It's not to say that moral principles themselves are subjective, only that desires are informative of who moral principles and rules apply to. I give an argument here about the marginal cases—showing that it's plausible to think of which creatures have moral obligations based on which ones have moral desires.

The second important feature of morality that I address is our ability to condemn people who don't follow the moral law. When someone acts badly, it seems important to be able to criticise their actions for failing to

meet that objective moral standard. Here I respond in three parts. I begin by arguing that most agents will still be susceptible to the kind of moral and rational criticism that we expect. Next, for when that isn't the case, I argue firstly that we have better kinds of criticism to offer, and then that certain types of criticism in these cases would be inappropriate anyway.

1. Hypothetical Imperatives and the Desire-Based View of Normativity

Hypothetical imperatives are, as I will understand them, very simple. They're (1) imperatives (statements that prescribe what to do) that are (2) hypothetical (conditional on something). The thing that they're conditional on is the agent's current set of desires.[1] They take this form:

$$\text{If A desires X, then A ought to } \varphi^2$$

Where A is an agent, X the state of affairs that they desire to come about and φ is an act that might[3] bring about that state of affairs. Take Bren as an example, and we have the following:

If Bren wants to help others, then Bren ought to go to the protest.

Imperatives of this form can be very wide-ranging. They can include agents' better and worse desires, and a number of different actions that might bring about those ends. Not all imperatives of this kind will be very important, but the connection with desire is what gets the normativity going.

A desire-based view of practical normativity, then, is one where everything that an agent ought to do takes this same form.[4] There is some desire, either explicit or not, in the foreground or the background, that plays an essential role in the explanation of why she ought to perform the action.

---

[1] Kant (2012) is specifically interested in what we will, rather than what we desire. Since I'm interested in the wider concept (desires construed broadly) I'll stick with that. This is also a move that Wedgwood (2011) and Smith (2004) make, according to Kolodny and Brunero: "Some suggest that this focus, on intentions and beliefs about necessary means, inspired by Kant's initial discussion of hypothetical imperatives (…) is overly narrow (…). Not simply intentions, but also desires, should be considered (…)". (Kolodny and Brunero 2020)

[2] In normal language it's often the case that the former, conditional, part of the imperative is left implicit. Finlay explains this in detail in (2014, 146-175).

[3] I say 'might' to remain neutral on the relevance of the agent's epistemic perspective. It might be the case that hypothetical imperatives apply to an agent if the action will bring about something that the agent desires, or it might be the case that they apply to an agent if the agent believes that the action would bring about something they desire.

[4] Or that they have *reason* to do, depending on the specific theory.

In the contemporary literature, a lot of the discussion of desire-based normativity is framed specifically in terms of normative *reasons*. According to this view, an agent has a practical normative reason to act if there is an appropriate connection between that reason and their desires. It can be helpful to contrast this with 'motivational' or 'explanatory' reasons, which look to give *explanations* for why someone acted. 'I fell over because I was clumsy' gives an explanatory reason, 'I fell over because I wanted sympathy' gives a motivational one.

Normative reasons do more than just explain actions or explain an agent's motivations—they *justify* those actions. Alvarez, who explains the distinction well, says, "it seems clear that reasons can have normative force. By that I mean that reasons can make something right—not necessarily morally right, but right in some respect" (Alvarez 2009, 182). That's what makes a reason a normative one, rather than an explanatory or motivating one. It's not just that we want to understand why I fell over mechanically or in terms of what mental states featured in my reasoning— normative reasons cover the *why* as well as the *how*.

There are several reasons why a desire-based view of reasons (or practical normativity more generally) is appealing. The source of the normativity's justification is clear: it comes from the desire, we see in exactly what respects these reasons are making something 'right'. We know why Bren ought to go to the protest, and it's because of her desire to stand up for justice. It also shows us why the specific agent in question ought to act, and under which conditions the imperative stops applying. It's not metaphysically weird.[5] It gives us a link between the action and the psychology of the agent, narrows the list of things she ought to do down to things that she is (in some sense) capable of doing (Williams 1981; Goldman 2009; Markovits 2014). It gives us a way to distinguish between normative behaviour like reasoning (and appealing to an agent's desires) with non-normative behaviour like restraining (Williams 1981; Manne 2014). All of these arguments are spelled out more carefully elsewhere. My own task will be to demonstrate the account's compatibility with important features of morality that are usually thought to be necessary sacrifices for such views. In doing so I will give the reader even more

---

[5] See Mackie (1977) and Joyce (2001). Bedke (2010) is an example of an opposing view here, he disagrees that desire-based theories of normativity are any less weird than objective theories. He argues that the desire-based theorist and the objective theorist have a "sincere disagreement" (see Bedke 2010, 50) when they discuss what counts as a normative reason, and therefore they can't mean such completely different things by the term that, for the desire-based theorist, it's conceptually necessary that all normative reasons must have a relation to the ends of the agent. I disagree here, as other desire-based theorists might – I think this part of the definition of a normative reason is, after all, the crux of much of this debate.

reason to find the desire-based view persuasive. Or at least, take away a reason they might've had to rule it out.

Before moving on to describe my opponents' views, I will briefly pause for some clarifications on what I mean when I talk about desire.

1.1 On Desires

One key part of my understanding of hypothetical imperatives is the broad understanding of desire mentioned above. I take desires to be a certain kind of pro-attitude that can manifest in a wide variety of ways, but there's plenty about desire that I want to remain neutral on. I will say that I don't take 'desires' to only mean desires that an agent is conscious of at a given time, or desires that the agent feels particularly strongly. Even our ordinary idea of desire, after reflection, includes a far greater range of desires than that.

Foot gives the following examples when she discusses hypothetical imperatives:

> Sometimes what a man should do depends on his passing inclinations, as when he wants his coffee hot and should warm the jug. Sometimes it depends on some long-term project, when the feelings and inclinations of the moment are irrelevant. If one wants to be a respectable philosopher one should get up in the mornings and do some work, though just at that moment when one should do it the thought of being a respectable philosopher leaves one cold. (Foot 1972, 306).

When I wake up in the morning and feel only the intense desire to stay in bed, that doesn't mean that all of my other desires have gone away. I still have projects that I desire to continue, people I care about and desire to do well, etc. I still want to become (or continue to be) a respectable philosopher, and I still want to get up in the mornings and do some work, even when those desires aren't reflected in my current phenomenology.

As well as the desires that I'm conscious of at a certain time, I also have what Pettit and Smith call 'background desires' (Pettit and Smith 1990), that is, desires that feature in the background of my thought rather than the foreground. I also want to include what Tim Schroeder calls "standing desires" (Schroeder 2017), desires that do not necessarily play a "role in one's psyche" (ibid.) at a certain time. After all, we sometimes act to

overcome the desires we *feel* most strongly. What moves us to do this are our other desires.[6]

The broad understanding of desire used in this paper is also compatible with a number of different theories of desire itself: that is, theories about what desire actually is. Arpaly and Schroeder, for example, argue that to have an intrinsic desire for P "is to constitute P as a reward", and to desire not-P is to constitute it as a punishment (Arpaly and Schroeder 2013, 127). There are also a variety of theories in which desire is a particular disposition, such as a disposition to act in a way to bring P about, a disposition to believe that P is good, or a disposition to *feel* attraction towards the prospect of P.[7] This paper argues that desire—whatever that turns out to be—can be a necessary feature of practical normativity, without that necessary relationship ruling out moral realism.[8] The only thing I require of a theory of desire is that it isn't too narrow, and that includes the wide range of phenomena listed above: from the background desires and 'passing inclinations' to the full-blown projects and passions.

I also take it that, for most of us, at least some of our desires will be moral desires. Desires to change the world for the better, to help our friends, to feed the hungry. Desires that lead us to doing good actions. The details of what might make a desire a 'moral' one might rely, to an extent, on which moral theory turns out to be true. It might be the case, for example, that our moral desires are those altruistic ones which compel us to maximise good for others, or to honour and respect the people we do meet. It might be the case that they are those which compel us to exhibit the right virtues, to be kind, generous, and forgiving.

But this doesn't mean that we would need to know the right moral theory for our desires to be moral ones. It might turn out that Marla's desire to help a hungry woman on the street is a moral desire even though she doesn't know the specifics of why it's a good thing to do, she just knows that she wants to help the woman.

---

[6] Another example of the kinds of broad range of desires I'm after is when Williams (1981, 105) refers to a subject's "motivational set". He includes in this set "dispositions of evaluation, patterns of emotional reaction, personal loyalties and various projects (…) embodying commitments of the agent". I take it that many of these will tend to coincide with what we desire.

[7] The former two theories here are discussed, along with a number of others, in Schroeder (2017). The feeling-attraction theory is argued in Smithies and Weiss (2019).

[8] It might be the case that some accounts of desire are better placed to explain *why* desires provide the normative force that they do. This might be the case with accounts that explain desire in terms of feeling attraction, for example, such as the accounts you might find in Chang (2004) or Smithies and Weiss (2019). I'm grateful to an anonymous referee for their insights here.

However it is that the distinction is cut, I take it that some of an agent's desires can be moral ones, and others not.[9] And if an agent can have moral desires, then perhaps they can also have moral reasons and moral oughts that apply to them in virtue of those desires. This paper defends the idea that these moral oughts can *apply* to agents in virtue of the agent's desires, and that such dependence on desire is compatible with those oughts being genuinely moral oughts.

Next, I will turn to understanding the rival accounts according to which such compatibility isn't possible.

2. The Opponents

As I mentioned in the introduction, this paper has two main forms of opposition. Firstly, it will target those who agree that normativity must be connected to desire but go on to suppose this gives them reason to be sceptical of moral realism. Secondly, I will cover those who see that same dilemma, and, if forced to choose, would rather hold on to moral realism than be tempted by an account of normativity that makes imperatives contingent on desires. I hope to give people in both camps a reason to look again at whether they need to make that choice at all.[10]

A common route to go down, for people who are persuaded by desire-based theories of normativity, is error theory about morality. Joyce (2001) makes one such argument, although similar ones are made by other error theorists such as Olson (2014) and Mackie (1997). Joyce's argument can best be explained by looking at a distinction between two kinds of categorical imperative:

2.1 Weakly Categorical Imperatives

Firstly, some imperatives apply to agents in a 'weakly' categorical sense. This amounts to not much more than being a description of a set of rules, or perhaps a description of what the speaker would prefer for the agent to do. Joyce gives an example of the rules of gladiatorial combat (such as not to throw sand in your opponent's eyes), and an unwilling gladiator called Celadus.

---

[9] Some good discussion on this topic – the 'moral worth' of an action, and its relation to desire, can be found in e.g. Arpaly (2002) and Sliwa (2016).

[10] This compatibility with moral realism, and with certain important features of morality, is what makes my project distinct from others who want to defend desire-based views of morality but who don't see this as being compatible with such features, such as Harman's (1975) moral relativism or Street's (2008) constructivism. Instead of arguing for a moral relativism, I'm arguing for a relativism about normativity, and objectivism about morality.

He says,

> When we say that Celadus ought not throw sand in his
> opponent's eyes, this is a weak categorical imperative. A
> Roman spectator—someone who heartily endorses gladiatorial
> combat and all its rules—will not retract her judgment
> "Celadus ought not throw sand" just because it is pointed out
> to her that Celadus wants to throw sand, and throwing sand is
> the best means of Celadus satisfying his own desires. (Joyce
> 2001, 36)

He also refers to these weakly categorical kinds of rules as the "non-
evaporating" kind: the kind of rule that doesn't evaporate just because the
agent has none of the relevant desires.

It's also the same kind of 'categorical' that Foot (1972) addressed when
she makes an analogy between morality and rules of etiquette. Just because
someone doesn't care about the rules of etiquette, that doesn't mean that
the rules don't apply to them. But these kinds of weakly categorical
imperatives don't—on their own—seem to come with any "practical
oomph" (Joyce 2006, 63) of normativity. They seem to be *de*scriptive more
than *pre*scriptive. This isn't what my opponents are looking for.

2.2   Strongly Categorical Imperatives

The alternative for my opponent is to say that some imperatives apply to
agents in a 'strongly' categorical sense. This is the sense that goes further
than weakly categorical imperatives, further than just describing rules that
apply to certain agents. Strongly categorical imperatives apply regardless
of an agent's desires (just like weakly categorical imperatives do), but they
also come with the oomph of normativity that's missing from weakly
categorical imperatives. Joyce doesn't say much more to explain what that
extra step looks like, and for good reason—he doesn't think it exists. Foot
(1972, 314), too, talks about it as a feature that's "missing".

According to the error theorists, morality *needs* to consist of imperatives
that apply in this strongly categorical sense. Morality needs something
stronger than weakly categorical imperatives can supply, and that
something is a "non-negotiable" (Joyce 2001, 8) part of our moral
discourse. That is, without that necessary feature, our moral discourse turns
out to be in error. Because the error theorists think that this extra ingredient
does not exist, that we can *only* understand normativity as being related to

an agent's desires, they also think there can be no such thing as morality after all—our moral discourse is all predicated on a mistake.[11]

Unfortunately, most moral realists tend to agree that morality must be strongly categorical in Joyce's sense. In a recent paper on error theory, for example, Lofitis says "Pretty much everyone accepts the unconditionality of moral requirements (…). Crucially, both error theorists and moralists accept the reason-supplying force of morality" (Lofitis 2020, 40). Shafer-Landau argues against desire-based theories of normativity in 'Moral Realism'.[12] He, like the error theorists, seems to think that such theories do not make room for some key concepts that are important to moral realism. It's these key concepts I'll turn to now, as I begin my defence of this hitherto unpopular view.

3. Two Features of Morality

3.1 Authority (and Inescapability)

The first feature of morality to discuss is its inescapable nature.[13] This is the idea that our moral imperatives have an authority over us: an authority that we can't escape from or choose to ignore.

Before I begin, I'll briefly say something about what I mean by imperatives a) having authority over us and b) being inescapable. I take these to be similar ideas, as they aim to explain a way in which the imperative is important, and something that we cannot opt out from. Imperatives have authority over us when they come from an (authoritative, important) source, and they are inescapable when they apply to us whether we like it or not. I aim to show that both of these ideas can be the case for hypothetical imperatives, and I'll flesh out more about what I mean as I show why this is so.

---

[11] I should note that not everyone who finds this part of morality puzzling is most commonly known as an error theorist. Williams, for example, argues for a desire based theory of normativity in Williams (1981; 1995) and then against the strongly categorical nature of morality in Williams (2011).

[12] See Shafer-Landau (2003, 165-167) in particular. Another moral realist, (Brink 1989), is at least sympathetic to the possibility of questioning the strongly categorical nature of moral imperatives, even though he doesn't go down that route himself.

[13] Williams (2011), for example, discusses morality's inescapability in Williams. Joyce also (2001, 51) talks about hypothetical imperatives not being inescapable. Sinclair in his (2016) talks of the inescapability of moral reasons.

First Mistake: Escapable Desires

Firstly, agents can no more 'escape' moral hypothetical imperatives than they can 'escape' their moral desires. I can't stop wanting to be a good person when that want is inconvenient for me. That's why we feel pangs of guilt when we do (or we're tempted to do) something we know is wrong, and why we'll often at least *consider* doing what's right, even when the less morally good option is so appealing.

Suppose that Bren has to choose between going to a protest for a good cause and staying in to have a quiet day at home. Suppose as well that the right thing to do is to go to the protest, and that she knows it, but the better thing *for her* and her happiness would be to stay at home. Despite this, she still wants to help others, and she would best do this by going to the protest. This desire isn't something that Bren can opt out of. It might be raining outside and her flat is warm and cosy. Her cat is asleep on her lap and there are games that she wants to play. The protest is a crowded bus journey away, and the weather is cold and uninviting. All of these facts might be running through her mind as she decides what to do, and the temptation to stay at home can be incredibly strong. But the reason that it feels like a difficult decision at all is *because* her desire to help others is still there, the moral imperative still hangs over her. And if she's made her decision and she wants to stop feeling bad about it, she might have to trick herself into not thinking about it any longer, or kid herself that it wasn't *really* the right thing to do after all. Perhaps, she thinks to herself, her presence wouldn't have made a difference anyway. And it really wouldn't be right to move the cat.

Not all of our decision-making will be like this, but the example's familiarity is indicative of the persistence of some of our desires. In many more cases, our moral desires will still affect us but be less obvious to us. Think back to the breadth of desire—including those that the agent isn't conscious of at the time. For many of us, the desire to be good is not a whim or a short-term desire that just pops into our heads in certain situations and then goes away again; it's a *long-term* preference, a standing desire. We tend to want to be good people throughout our adult lives, and this can persist even when the alternative actions seem much more tempting for one reason or another. It's something many of us want for ourselves overall, when we get the chance to sit back and think about what's really important. We shouldn't be surprised, then, that such a desire can sometimes fail to translate proportionately into a strong motivation, or even into a noticeable feeling. But when the desire isn't manifesting in those ways, it doesn't mean that the desire isn't still there, or that the moral

imperatives would cease to have a hold on us.[14] Bren still ought to go to the protest, and the tired philosopher still ought to get out of bed and do some work.

It's important that a theory of normativity be able to explain what it's like to feel conflicted about our obligations, and to account for how some of our strongest obligations can be the ones that seem to apply to us *despite* the fact that we really don't want to fulfil them. But having an agent's imperatives be contingent on her desires doesn't mean that there'll never be conflict and struggle, that there'll never be tough situations, or that there'll never be times that she is compelled by moral imperatives to act against what her strongest desires *seem* to be. With morality as a system of hypothetical imperatives, the struggle is (still) real.

Second Mistake: A Relative Morality

A related point is one about whether there can be objective moral standards, which exist independently of ourselves. The existence of such principles might be the kind of thing that concerns some people when they think about the 'authority' of morality—the fact that we must be held accountable to a standard beyond what we happen to desire.[15]

The existence of strongly categorical imperatives isn't necessary for an objective moral standard itself to still exist. Morality can still form an independent standard without everyone always having reason to follow it, without moral imperatives applying to every creature out there. Only people with the right moral desires will have moral imperatives apply to them, but there can still be an objective truth about which desires are moral. An analogy here might be with other weakly categorical rules, like the law. The law of a country (we can suppose) is a defined set of rules that exist independently of that country's citizens and their desires. This doesn't mean that everyone has a reason to follow all of the laws of that country. Some laws might be unjust or irrelevant (the former being something less

---

[14] It might be an additional worry to some that our moral desires, although they might be continually present, might often not be our *strongest* desires, and so the moral course of action might rarely be what our desires would point us towards overall. I don't see this as a particularly big problem for the desire-based normative theorist. They could respond in three ways: firstly, by denying that desire strength is proportional to normative strength (instead it might just be the case that some desire is needed to get normativity off the ground at all). Secondly, they might gesture to the fact that subjective strength of desire is going to be complicated – and not necessarily correlated to how strong a desire *feels* to an agent. The kinds of desires that moral desires are (persistent, and the kind that we're likely to endorse) might mean that they tend to be quite strong after all, and have a lot of normative weight. Thirdly, they can bite the bullet, and say that it just is the case that we won't always/often have the *most* reason to do the morally best action.

[15] "Anyone who offers an account of the morality of right and wrong is bound to be asked whether he is claiming that there are 'universal' moral principles" (Scanlon 2000, 328).

likely to be the case in morality), or some people might have cares, projects or other reasons that prevent the law from being reason-giving. Perhaps most significantly, not all people are even going to be citizens of that country.

Morality is like the law in that way. It can be an objective standard and authority, one that exists whether or not there are any agents in the world to whom the moral obligations apply. For example, it might be the case that moral goodness consists in agents exemplifying the virtues of compassion and wisdom, or it might be when people work to improve the well-being of others. And if there are agents who *are* capable in at least some sense[16] of doing these things, who have the right desires, then those are the people who this objective law applies to, the people who have moral obligations. Fortunately, these are all real desires that people have. The normativity of morality is explained.

There's one more important point I want to make before I move on. An account of hypothetical imperatives doesn't imply that moral principles change with the desires of the agent, but rather the extent to which the objective principles apply to the agent can change with the agent's desires. Now is a good time to argue why that's plausible, and I'll do so here making particular note of marginal cases.

In the legal analogy there are people for whom the laws don't apply. The law doesn't apply to babies, to animals, or to people who live outside of its range, for example. The same applies in the case of the *moral* law. There are lots of entities without moral obligations: definitely rocks and trees, for example. Animals probably don't have moral obligations either, at least certainly not *most* animals. Babies also don't seem to start out with any, but will usually get closer and closer to being moral agents the older they get and the more they develop (or you might think that they acquire them all at once at a certain point in their lives; either way they tend to start out with none and end up with some).[17] An account of moral 'oughts' that connects them necessarily to the agents' desires gives us an explanation as to why this is: moral oughts apply to entities if/when those entities start to

---

[16] Given the broad understanding of desire here, I take it that an agent is (in at least some sense) not able to act without having a desire to do so.

[17] This is a different question to whether babies have moral status. Presumably they'll be worthy of our moral attention long before they start having moral obligations of their own.

have moral desires.[18] And that answer seems to track the marginal cases in the right way.[19] Here I'll go through some of them.

Inanimate objects don't have desires, and they don't have moral obligations. That one's fairly simple, but also not particularly informative. There are lots of other things, after all, that inanimate objects don't share with humans and which might explain why they don't have moral obligations. They also don't act, for example. Animals, then, might have desires, but they're unlikely to have the kinds of desires that would satisfy an ethical theory's conditions for being *moral* desires. Animals don't generally desire to be good in the way that we take to be morally relevant. But what if they did? If there are intelligent animals out there who form genuine friendships or loving relationships, who are able to understand some concept of goodness that's recognisably moral and to then intrinsically desire to act in good ways or bring goodness about, then it seems like there would be a good case for those animals being subject to moral oughts. This all depends, of course, on our understanding of the psychology of animals and what our preferred ethical theories take to be the right kinds of moral desires. But so far, having moral desires is a plausible criterion for moral agency.

The case might be easier to imagine with children, who regularly *do* turn into functioning moral agents. Again, according to my account, they do so not when they want to tell the truth or resist drawing on the walls only in order to avoid getting into trouble (because that isn't likely to qualify as a moral desire), but when they start wanting to do so just because they want to do what's right, or they understand the harms involved and want to avoid them, or because they see truth as having value (depending, of course, on the details of what you think a moral desire looks like).

---

[18] A better laid out explanation of this can be found in Arpaly and Schroeder (2013). Aristotle (2004), too, spoke about how taking pleasure and pain in the right things is something that can be brought about through moral education. This is mentioned in Homiak (2016). Mark Schroeder agrees, saying that "Virtue (…) involves desiring the right things, and to the right degree" (2004, 177). For more discussion of when an agent might qualify for moral obligations see, for example, Alvarez and Littlejohn (2017) who talk about a distorted capacity for moral thinking. This is compatible with my view that it's dependent on an agent having moral desires, since a lack of moral desires might do just that. This is similarly the case with Rosen's (2004) suggestion of brain anomalies, or of being badly taught.

[19] It's worth noting that even if agents do reliably qualify for moral obligations at the same time that they get moral desires, that doesn't necessarily mean on its own that the moral desires caused them to become moral agents. It could be, for example, that both arrive at the same time because of a third factor that causes them—such as the agent being able to recognise moral reasons. I don't mean for this correlation to be definitive proof of desire-based theories of normativity (and I want to thank an anonymous referee for this point, which I had missed on my own). I do, though, take it to be a point in the theory's favour over accounts that don't have a way to explain moral agency as easily. I also take it to be a point of defence for the theory, against those who think that desire-based theories of normativity don't do a good job of tracking who it is that moral obligations apply to.

Until people have these desires (if they ever do acquire them), when we tell them that they *ought* to do what's right, and talk to them about their moral reasons, it doesn't seem like we're using the terms in a seriously *morally normative* sense, as much as it seems like we're training them. The imperatives seem only to be weakly categorical. We're showing them the kinds of things that we want them to desire, that we want them to take to be important. We're describing the moral law as we see it and we keep doing it until they can see it for themselves. Until then, we'll make do with the fact that they still ought to avoid drawing on the walls because it *will* get them into trouble when they're caught.

To complete the picture, we should think about what happens when humans really never do acquire the right kinds of moral desires. Firstly, it's worth saying that these sorts of people are very rare. They're who we might find ourselves describing as amoralists. They're people who are set up in such a way that they don't have the right kinds of moral desires, and therefore cannot do what's good for the right kinds of reasons. My account here says that these people are not subject to moral oughts or moral reasons, and I'm more than happy to agree with that. (They might become subject to them at a later date. Perhaps with a good education or a good friendship, they might change.)

Before moving on I want to emphasise that this account is not trying to be particularly revisionary. In fact, I'm arguing for the opposite: that this account of morality, including of who is subject to moral oughts and who isn't, is in keeping with what we should expect. In the same way that it wouldn't be fair to view (most) very young children as being subject to normative moral requirements, the same is true of the rare cases of agents with no moral desires. The desire-based account is stronger for its way of explaining these marginal cases.

Anyone who claims that morality applies categorically will have some caveats, but there's just some disagreement about what those caveats are. Presumably for now it only applies to things that are alive. Perhaps, for some, it applies to creatures with a certain ability to reason or a rational capacity. For the desire-based account, it's creatures with certain desires. I've shown here that these accounts aren't so different.

3.2   Moral Condemnation

Another important feature of the moral system is moral condemnation. I'll begin by explaining why it's taken to be an important part of moral discourse, and why my opponents might worry that such criticism isn't possible on my account. This time I'll respond to the worry in three parts.

Usually when we see an immoral action we condemn it, we think it worthy of our criticism. There's a moral standard (that can be universal, as just discussed) and if an agent in question fails to meet that standard then we can criticise them on that basis. It seems like something more than just personal disapproval of that agent's actions, but a more substantive criticism that's supported by a real morality.[20] This kind of criticism might seem like an important part of moral discourse.

Moral realists might worry about the status of moral condemnation if we accept a moral system that's only made up of weakly categorical imperatives, instead of strongly categorical imperatives. The latter, after all, are those that apply to agents regardless of what they desire, *and* that come with some kind of normative force. If moral 'oughts' only apply in a strongly categorical way to people who have moral desires, then my opponent might worry that we cannot morally criticise exactly the people who are the most in need of our criticism.

Here I'll begin by conceding a point. For the agent who has no desires at all to be a good person, it might just not be the case that they ought to do what's right. There are a range of criticisms unavailable to us about such an agent: she's not being irrational, for example, and she's not failing to do what she ought to do. We cannot criticise them in the kind of way that Kiesewetter seems to want to, when he talks about moral criticism and says,

> Criticising someone involves more than the judgment that the criticised person has violated some standard; it also involves the judgment that the standard is authoritative for her. […] this means that the person has decisive reason to conform to this standard. (Kiesewetter 2017, 25)

He goes on,

> It seems blatantly incoherent to maintain a criticism while accepting that the person criticized had sufficient reasons for what she is criticized for. (Kiesewetter 2017, 29)

This objection also comes up against desire-based theories of what we have reason to do. Williams, arguing in favour of such a theory ('reasons

---

[20] Smith, Lewis and Johnston, for example, describe a "panic" at the idea that there isn't an objective rationale for morality because, for one thing, our disapproval at each other's moral views wouldn't be the same kind of serious thing we took it to be (Smith, Lewis and Johnston 1989, 103-104).

internalism') discusses such a case, of a man who has no desire to do the right thing. Williams says,

> There are many things I can say about or to this man: that he is ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal and many other disadvantageous things. (…) There is one specific thing that the external reasons theorist wants me to say, that the man has a reason to be nicer. (Williams 1995, 39)

But I still think that accounts of desire-based normativity have enough tools to deal with this problem, and I'll show how in three ways.

Response One: Most Agents Will Still have Moral Oughts

My first response is this: agents have a variety of desires, and what they desire most overall will often be different to what they feel most strongly at any given time. This has become a familiar theme in this paper: I don't stop wanting to be a productive philosopher when I'm in bed in the mornings, and I don't stop wanting to do what's right when I'm faced with an incredibly difficult moral decision. Because of this, we *can* often criticise people for failing to do what's right, and we can do so on the grounds that they're not correctly adhering to their desires or to what they ought to do. When an agent fails to do something good we can criticise them because they're too busy paying attention to their shorter term desires over their longer term ones, choosing the easy options over those that will help them fulfil what they want the most overall.

Brink makes an important point here, following Hume. He says,

> If, for example, sympathy is, as Hume held, a deeply seated and widely shared psychological trait, then, as a matter of contingent (but "deep") psychological fact, the vast majority of people will have at least some desire to comply with what they perceive to be their moral obligations, even with those other-regarding moral obligations. Moral motivation, on such a view, can be widespread and predictable, even if it is neither necessary, nor universal, nor overriding. (Brink 1989, 49)

It seems, fortunately, like most people *do* have moral desires, even if we're often bad at acting on them, bad at prioritising them, or bad at seeing *how* to act on them. It seems like it just happens to be a fact that most people, therefore, ought to act morally, even if what they ought to do is contingent on what they desire.

16

At other times, the reason why people fail to do what's good will be down to an epistemic mistake of some kind. I might have moral desires but be really bad at recognising good actions when I see them, for example. Here it would still be the case that I ought to do what's *actually* right, because, presumably, I'd still want to do *what was right* and I'd be failing to do it. If someone mistakenly thinks that the right thing to do is to catcall the woman walking past ("I'm such a thoughtful person," they think, "I'll just give her a quick boost to her self-esteem") then it's still open to us to criticise them for not getting the facts right.

Response Two: Criticism on Other Grounds

Perhaps that was a bit too optimistic an answer, to declare that we do have strong moral desires most of the time. But luckily I have two more responses in stock. Secondly, then, even when an agent's desires, correctly weighed, *don't* give them the most reason to do what's morally right, we can still criticise them on other grounds. Indeed, I'll argue that on these occasions it would be far more appropriate to criticise them on these other grounds, rather than on the grounds of irrationality.

As a reminder, we can't criticise these agents for not following their reasons correctly, for being irrational or for failing to do what they (normatively) ought to do. (We can, of course, still say that they 'failed to do what they ought to do', but in a way that acts more like a description of moral rules or of our own preferences than a normative prescription—the rules that apply to them in a 'weakly categorical' sense.) But there are other significant criticisms that we can make. Perhaps most notably, we can criticise them for not having the right (moral) desires in the first place.[21] We can call them cruel, thoughtless, callous, selfish, mean-spirited. The most relevant criticisms that we tend to use in these times seem like they're descriptions of the agent's psychological state: criticisms of their priorities, their desires.

Suppose we want Marlene to join us at the protest today, but she doesn't care about the political issue that's riled us up. Morally speaking, it seems to be the case that she has an obligation to get involved—she's a member of a privileged group and there is a clear injustice that she could help to prevent by attending. But, on balance, she just doesn't want to help—and certainly not more than she wants to stay at home and re-organise her bookshelf. She has no good excuses or mitigating factors: she's not socially anxious, she's not tired, and she would be in a good position to make a difference. Just because we can't criticise her for being irrational, we can

---

[21] Arpaly and Schroeder (2013) argue that to be good (and virtuous) is to have the right kinds of desires.

still criticise her in a variety of more important ways. Marlene is being selfish, she's failing to adequately care about the plight of those less fortunate than her, and she doesn't have a strong enough desire to do what's right. It doesn't matter that she's not failing, by her own lights and according to her own desires, but it matters that she's failing to meet *our* standards, and failing to meet the objective moral standards that are important *to us*.

This doesn't seem to be an uncommon view; it's what we do when we call people callous, rude, or selfish. Brink makes this point,

> (…) Moral requirements would still apply to agents independently of their contingent and variable desires, even if they would not provide agents with reasons for action independently of their desires. Thus, we could still charge people who violate their moral obligations with immorality, even if we could not always charge them with irrationality. (Brink 1989, 75)

Foot, too, makes a similar point, when she says "The fact is that the man who rejects morality because he sees no reason to obey its rules can be convicted of villainy but not inconsistency" (Foot 1972, 310). Being convicted of villainy seems like plenty, for the limited number of situations in which that's all we can do.

This response is also relevant to other forms of criticism that might *seem* at first to be criticisms of irrationality. Take 'thoughtlessness' for an example. My opponent might worry that on my view it's difficult to criticise moral agents as being thoughtless, because it's a criticism that seems to aim at agents who have failed to notice certain things, failed to think or deliberate properly. The one kind of criticism that my system *can't* account for is to criticise people on the basis of failing to follow their own reasons when they have none of the relevant desires. But to fail to think and attend properly to certain things can be a result of not having moral desires. Arpaly and Schroeder (2014, 227), for example, list four ways that desire can affect cognition other than through directly affecting action.

1.   Through involuntary shifts in attention
2.   Through changing dispositions to learn and recall
3.   Through changes in subjective confidence
4.   Through distortion by emotions and wishes

Because our desires affect the way we learn and recall things, our confidence, our attention, etc. they affect what we do beyond just affecting

what choices we consciously make. Criticising someone on the basis of having the wrong desires encompasses a lot.

Perhaps this response just puts the problem one step back. How do we criticise an agent for their desires, when it's not the case that they ought to have different desires? But here we can just put our feet down. We can criticise others without needing to criticise them by their own lights, in a way that is authoritative for them (as Kiesewetter thought was important). We can just criticise them, full stop. And there are still standards by which these agents can be wrong, other than the standards of rationality. The immoral agent with no moral desires is doing something wrong by moral standards, even though she's not doing something irrational. We can still use weakly categorical oughts when we talk about them, it's just that this is describing a set of rules (such as the moral law) that they're failing to meet, rather than saying what they normatively ought to do.

It might be the case that some people think of this as something more akin to a descriptive claim, rather than criticism as such. But this seems to have too narrow a view of what criticism is. Think of other things I can criticise: I can criticise a bird for being too loud, an artwork for being boring, or a tree for being ugly. I can criticise a person for being bad at sports, or a storm for causing damage to the trees in my local park. It seems unnatural to say that I am merely describing these things, when what I am really trying to do is to be critical, to say something negative about them. Even when none of the targets of my criticism ever had a reason to be otherwise, and don't care about what I have to say—the words can still serve the purpose of expressing my own disapproval. I think that our moral criticisms are often closer to this kind of criticism, and that's ok.

Furthermore, our criticisms can often have an important effect even when we're not using them to persuade people to follow their moral obligations. For example, suppose I see someone throwing litter on the street and I call them out for it. This person might not care at all about clean streets or nature or keeping the local area beautiful for people and animals to enjoy, etc. But my criticism might serve to motivate them in other ways. Perhaps by criticising them, I might appeal to reasons they *do* have—such as reasons to avoid looking bad in public. It's unlikely that this would be an appeal to a *moral* desire, and that they would therefore be acting moral as a result, but my moral criticism here can still serve a purpose of getting them to pick up their litter. Perhaps it might even lead to a change in attitude down the line that eventually leads them to have the right moral desires after all. If enough people show you what they value, perhaps that can give you reason to consider taking those values on yourself.

Response Three: When the Criticism is Inappropriate

I have one final defence of the ability to morally criticise people under a system of hypothetical imperatives. I've already argued that we can criticise agents with no (or not enough) moral desires in a number of ways, and that we can still criticise people in a lot of cases for failing to do what they ought to do. Finally, I'll say that the people for whom the last category doesn't apply—the people without the correct moral desires—are simply *not* the people who normative moral imperatives *should* apply to. This is something that I argued for more in the previous section, but is also relevant here.

When we criticise a very young child for causing someone pain, we'd be *wrong* to criticise them as being irrational or for failing to follow their own reasons. It's not the case that the moral law already applied to them and they just failed to act in accordance with it. After all, they don't have the moral maturity yet to have moral reasons. And it's plausible to say, I think, that this is because they don't have the right kinds of moral desires yet. In criticising people without adequate moral desires for being irrational we'd be making the same kind of mistake. Better to criticise them in a different way, and/or do our best to instill and encourage the right kinds of desires in them in the future.

Conclusion

The aim of this paper is to convince my opponents that we can have a plausible, familiar, and real moral system while still understanding moral imperatives as being hypothetical, rather than categorical. I did this by addressing two important features of morality—the two that might most have been in danger if we understood morality to be a system of hypothetical imperatives—and I showed how they could both be accounted for.

Although this paper should be of interest more generally to people who care about ethics, meta-ethics, and moral psychology, I also had two main opponents: firstly, some of my opponents are moral non-realists. That is, people who think that moral realism is implausible *because* moral imperatives cannot be strongly categorical. Secondly, my paper also targeted moral realists, who would rather give up the idea of morality as a system of hypothetical imperatives than morality itself. Instead, I hope to have shown, we can have our morality and eat it, too.

Acknowledgments

REFERENCES

Alvarez, Maria, and Clayton Littlejohn. 2017. "When Ignorance is No Excuse." In *Responsibility—The Epistemic Condition*, edited by Philip Robichaud and Jan Willem, Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198779667.003.0003.

Alvarez, Maria. 2009. "How Many Kinds of Reasons?" *Philosophical Explorations* 12 (2): 181-193. https://doi.org/10.1080/13869790902838514.

Aristotle. 2004. *The Metaphysics*. London: Penguin Books.

Arpaly, Nomy. 2002. "Moral Worth." *The Journal of Philosophy* 99 (5): 223-245. https://doi.org/10.1093/0195152042.003.0003.

Arpaly, Nomy, and Timothy Schroeder. 2013. *In Praise of Desire*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199348169.001.0001.

Bedke, Matthew. 2010. "Might All Normativity Be Queer?" *Australasian Journal of Philosophy* 88 (1): 41-58. https://doi.org/10.1080/00048400802636445.

Brink, David. 1989. *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press.

Chang, Ruth. 2004. "Can Desires Provide Reasons For Action." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz,* edited by R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, 56-90. Oxford: Oxford University Press.

Finlay, Stephen. 2014. *A Confusion of Tongues*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199347490.001.0001.

Foot, Philippa. 1972. "Morality as a System of Hypothetical Imperatives." *The Philosophical Review* 81: 305-316.

https://doi.org/10.1093/0199252866.003.0011.

Goldman, Alan. 2009. *Reasons from Within*. Oxford: Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780199576906.001.0001.

Harman, Gilbert. 1975. "Moral Relativism Defended." *Philosophical Review* 84 (1): 3-22. https://doi.org/10.2307/2184078.

Homiak, Marcia. 2016. "Moral Character." *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), edited by Edward N. Zalta. Available online at:
https://plato.stanford.edu/archives/fall2016/entries/moral-character/.

Joyce, Richard. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511487101.

Joyce, Richard. 2006. *The Evolution of Morality*. Cambridge, MA.: MIT Press. https://doi.org/10.7551/mitpress/2880.001.0001.

Kant, Immanuel. 2012. *Groundwork for the Metaphysics of Morals*. Cambridge: Cambridge University Press.

Kiesewetter, Benjamin. 2017. *The Normativity of Rationality*. Oxford: Oxford University Press.
https://doi.org/10.1093/oso/9780198754282.001.0001.

Kolodny, Niko, and John Brunero. 2020. "Instrumental Rationality." In *the Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), edited by Edward N. Zalta. Available online at: https://plato.stanford.edu/archives/spr2020/entries/rationality-instrumental/.

Lofitis, Kipros. 2020. "Standard and Alternative Error Theories about Moral Reasons." *Ratio* 33: 37-45.
https://doi.org/10.1111/rati.12252.

Mackie, John. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.

Manne, Kate. 2014. "Internalism about Reasons: Sad But True?" *Philosophical Studies* 167 (1): 89-117.
https://doi.org/10.1007/s11098-013-0234-3.

Markovits, Julia. 2014. *Moral Reason*. Oxford: Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780199567171.001.0001.

Olson, Jonas. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Pettit, Philip, and Michael Smith. 1990. "Backgrounding Desire." *Philosophical Review* 99 (4): 565-592.
https://doi.org/10.2307/2185617.

Rosen, Gideon. 2004. "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18 (1): 295-313.
https://doi.org/10.1111/j.1520-8583.2004.00030.x.

Scanlon, Thomas. 2000. *What We Owe to Each Other*. London: Harvard University Press.

Schroeder, Mark. 2004. *Slaves of the Passions.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199299508.001.0001.

Schroeder, Timothy. 2017. "Desire." In *Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), edited by Edward N. Zalta. Available online at: https://plato.stanford.edu/archives/sum2017/entries/desire/.

Shafer-Landau, Russ. 2003. *Moral Realism: A Defence. Oxford: Oxford University Press*. https://doi.org/10.1093/0199259755.001.0001.

Sinclair, Neil. 2016. "Reasons, Inescapability and Persuasion." *Philosophical Studies* 173 (10): 2823-2844. https://doi.org/10.1007/s11098-016-0639-x.

Sliwa, Paulina. 2016. "Moral Worth and Moral Knowledge." *Philosophy and Phenomenological Research* 93 (2): 393-418. https://doi.org/10.1111/phpr.12195.

Smith, Michael. 2004. "Instrumental Desires, Instrumental Rationality." *Proceedings of the Aristotelian Society* (Supplementary Volume) 78 (1): 93–109.

Smith, Michael, David Lewis, and Mark Johnston. 1989. "Dispositional Theories of Value." *Proceedings of the Aristotelian Society* 63: 89-174. https://doi.org/10.1093/aristoteliansupp/63.1.89.

Smithies, Declan, and Jeremy Weiss. 2019. "Affective Experience, Desire, and Reason for Action." *Analytic Philosophy* 60 (1): 27-54. https://doi.org/10.1111/phib.12144.

Street, Sharon. 2008. "Constructivism about Reasons." In *Oxford Studies in Metaethics* volume 3, edited by Russ Shafer-Landau, 207-247.

Wedgwood, Ralph. 2011. "Instrumental Rationality." In *Oxford Studies in Metaethics* volume 6, edited by Russ Shafer-Landau, 280-309.

Williams, Bernard. 1981. "Internal and External Reasons." In *Moral Luck.* Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139165860.009.

Williams, Bernard. 1995. "Internal Reasons and the Obscurity of Blame." In *Making Sense of Humanity.* Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511621246.004.

Williams, Bernard. 2011. *Ethics and the Limits of Philosophy.* Abingdon: Routledge. https://doi.org/10.4324/9780203969847.

SOMETHING NEGATIVE ABOUT TOTALITY FACTS

Andrea Raimondi[1]

[1] Thapar Institute of Engineering and Technology, Patiala, India

ABSTRACT

Armstrong famously argued in favour of introducing totality facts in our ontology. Contrary to fully negative (absence) facts, totality facts yield a theory of "moderate" or "partial" negativity, which allegedly provides an elegant solution to the truthmaking problem of negative claims and, at the same time, avoids postulating (many) first-order absences. Friends of totality facts argue that partial negativity is (i) tolerable vis-à-vis the Eleatic principle qua mark of the real, and (ii) achieves a significant advantage in terms of ontological parsimony. But are totality facts, which are partially negative, really more ontologically acceptable than fully negative facts? In this paper, we argue that, comparatively, the case for totality facts is weaker than commonly assumed and that, ultimately, the answer is negative.

Keywords: *totality facts; Armstrong; negative facts; truthmaking; causal powers.*

1.   Introduction

What determines the truth of negative claims is a daunting and lively problem in truthmaking metaphysics. It is significantly exacerbated if truthmaker maximalism is placed at the core of truthmaker theory (see Molnar 2000; Armstrong 2004; Merrick 2007; *inter alia*). The matter has elicited a range of responses. Some deny that there are negative truths (Mumford 2007), while others deny the need for truthmakers for the negation of atomic propositions (see Tallant 2010). For some others the quest for truthmakers is essential, but the project must not be pursued using negative facts; they do not deserve a place in our inventory of the world (see Rodriguez-Pereyra 2005, 31). As Varzi puts it, "we often talk *as though* there were such things, but deep down we may want our words to be interpreted in such a way as to avoid serious ontological commitment" (2006a, 132).[1]

Negative facts are not without defenders, though their supporters are partitioned among those, like Russell, who accept the existence of absence-negative facts, and others, like Armstrong, who favour limit-negative facts (Armstrong 2004, 53).[2] Assume, for example, the truth of "our roses are not blue". One candidate truthmaker is the absence-negative fact that [our roses lack blue]. Under the assumption that negative and positive facts have constituents, one is here called to bite the Meinongian bullet: these facts must have non-existent constituents. In fairness, it is a path that only a few braves have pursued. Alternatively, negative facts can be understood as primitive, *sui generis* facts. Still, as Molnar argues, "[t]his would be a particularly deep primitiveness, since negative states are not only a new kind of thing, they are a new kind of kind of thing" (2000, 77).

Armstrong famously advocates for another strategy where negativity is but one part of the solution to the truthmaking problem. Upon closer inspection, what makes our negative proposition true is what our roses are positively like, rather than what they lack: the conjunction of all their positive facts—e.g. that they are red, 1 mt. tall, standing in a white vase, and so on. Nevertheless, on the assumption that a truth-maker must necessitate a truth, we include in the conjunction a high-order, general fact: that the conjunction contains all the positive facts about our roses (see Armstrong 2007, 99). This particular fact is *negative,* since the high-order

---

[1] Others who resort to resources different from negative facts are Demos (1917), Heil (2000), Merrick (2007), Cameron (2008), Schaffer (2010), Beall (2000), *inter alia*. The attempt to avoid postulating negative entities is very often rooted in a general uneasiness famously described by Russell: "There is implanted in the human breast an almost unquenchable desire to find some way of avoiding the admission that negative facts are as ultimate as those that are positive" (1989, 217).

[2] Others who defend the view that absences are truthmakers for negative truths include Martin (1996) and Kukso (2006).

general fact that the first-order facts are *all* the facts just is the fact that there are *no more* facts. But, unlike absence-negative facts, this fact does something different than denoting an absence: it limits the number of first-order facts about roses. As Armstrong puts it, a limit-fact involves negation of a different sort: it does not involve a fact such that every rose is non-blue but one such that every rose is *different* from a blue rose, a "Plato's difference" (Armstrong 2004, 71). By Armstrong's admission, the solution does not dispel negativity altogether.[3] Limit-facts are as negative as absence-facts. However, limit-facts can only exist together with the positive facts that they limit because they denote a relation—the *-ally* or *totality* relation—that takes the aggregate of all the first-order states of affairs and totals them. As Armstrong puts it:

> [t]he aggregate stands in a highly specific relation to the instantiated property (…) of being an electron. That property may be said to total or to *all* that particular aggregate. (Armstrong 2004, 73)[4]

The limit-fact is thus the (negative) fact that no non-electron is a member of the aggregate and thus, in conjunction with the aggregate, serves as a suitable ground for truths about what there is not. The resulting relational fact (i.e. the *totality* fact) is only partially—rather than fully—negative because its constituents are both *positive*—viz. the big conjunction of positive facts—and *negative*—the limit fact (see Armstrong 1989, 92–97; Armstrong 2004, 54–70).[5]

If the game is to go-big or go-home, perhaps some might find it more attractive to side with Schaffer (2010) or Cameron (2008), who argue that we can banish negativity altogether by placing a different totality, the actual world itself, as a truthmaker for negative truths.[6] Nevertheless, for those who wish to defend negativity, totality facts seem *prima facie* more ontologically acceptable than fully negative facts. As Armstrong claims,

---

[3] As Armstrong puts it "George Molnar argued that there are four extremely plausible theses, which, however, cannot all be true: (i) The world is everything that exists, (ii) Everything that exists is positive, (iii) Some negative claims about the world are true (iv) Every true claim about the world is made true by something that exists. Molnar left the problem there. He had no solution. I submit with respect that in this situation the least evil is to reject (ii). The postulation of totality states of affairs, or at least one such state of affairs, is my way of doing that. Limits, if not absences, are ontological realities" (2004, 81-82).

[4] As Armstrong puts it, semi formally: *T* (aggregate of electrons, being an electron).

[5] This point is echoed by Tugby, who argues that "The higher-order totality fact can only exist as a constituent in a complex fact which involves positive first-order entities. This is why I take the totality fact approach to be a more moderate theory of negativity than one which accepts the existence of first order negative facts" (2017, 472).

[6] There are important differences between the two proposals. One above all is that, for Cameron, the (actual) world is constituted by the truthmakers for all positive truths (2008, 294). On the other hand, for Schaffer (2010) the world is the one fundamental substance that grounds all its proper parts.

they yield a "moderate" negativity that (i) is more tolerable vis-à-vis the Eleatic principle *qua* mark of the real, and that (ii) achieves a significant advantage in terms of ontological parsimony: "provided we allow ourselves general facts then no further negative facts are needed among our truthmakers" (Armstrong 2004, 54).

In this paper, we argue that, at scrutiny, the case for totality facts is weaker than commonly assumed and that, ultimately, totality facts are no more acceptable than fully negative facts. We are not the first to argue for this conclusion, which is one of the main take-home points of Molnar's seminal paper (2000), *inter alia*.[7] But Molnar's rejection of totality facts is based on the assumption that there is no real distinction between partial and full negativity and, hence, arguments against fully negative facts, viz. absence-negative, apply *mutatis mutandis* to totality facts.[8] We want to argue instead that even if the distinction between partial and full negativity is granted, totality facts are no more acceptable than fully negative facts: partial negativity is no better than full negativity.[9]

In § 2, we first argue that, based on Eleatic consideration, totality facts and fully negative facts are either equally ontologically unacceptable or equally ontologically legitimate. In § 3, we then proceed to present a truthmaking argument for the conclusion that the adoption of totality facts does not equate with an allegedly superior advantage in terms of ontological parsimony.

2.   Totality and Causation

The original argument from the Eleatic principle was initially proposed by Armstrong (2004, 76-77), but a more recent instalment is offered by Tugby (2017). As Tugby puts it:

> It is plausible that, unlike negative facts, totality facts satisfy Armstrong's own criterion for ontological commitment, which is known as the Eleatic Principle (…). Suppose I touch a live wire which is conveying a total current of 100 milliamperes. Suppose that I get severe burns as a result but survive because only total currents above 100 milliamperes would kill me. Here it seems that the relevant totality fact has a distinctive causal

---

[7] Another important paper in this respect is Barker and Jago (2012).
[8] For a similar argument see also Dodd (2007, 389).
[9] The claim that partial negativity is as objectionable as full negativity has been recently adumbrated by Ingram (2016, 1278).

> power: since the current was at 100 milliamperes and no more,
> I was burned rather than killed. (Tugby 2017, 473)

At least *prima facie*, Tugby's articulation of the Eleatic criterion closely aligns with the original Platonic formulation—"whatever has the ability to affect and be affected is a real being" (Plato, Sophist 247d-e). Accordingly, totality facts are *causal*—and hence real—in the sense of causally active or causally operative; *qua* causes, or *qua* bearer of causal powers, they are related to the effect by a relation of causal production.[10]

However, a second reading of the argument suggests that his formulation is also compatible with a second, "weaker" version, employed by Armstrong in his original argument (Armstrong 2004, 76-77; see also 1978, 25).[11] The causality of this "weak" principle is not a relation between cause and effect but instead between explanans and explanandum (see Molnar 2000, 78). Armstrong's favourite rendition of the Eleatic criterion attributes reality to something if it "makes a difference to the causal powers of something" (Armstrong 1997, 41). A property instantiation, for instance, can make a difference for the powers that a particular exhibits, but in itself does not cause the powers of the particular.[12] A property instantiation is instead a part of the causal explanation of how the particular comes to bear certain powers.[13]

Discussing a similar version of the principle, Molnar—building on Lipton (1993)—notices how it is often deployed when the explanation is contrastive, i.e. when the explanandum's relevance is compared with some other event (2000, 78-79). So, according to this version of the Eleatic principle, a totality fact is not causally active or operative but causally *relevant* because it explains why X came to have certain powers simpliciter, or why X occurred rather than Y—e.g. why "I was burned rather than killed" (Tugby 2017, 473). Both Tugby and Armstrong, independently, seem to agree in recognising the superiority of totality facts over negative facts vis-à-vis any formulation of the principle. The superiority depends on the failure of negative facts and the success of

---

[10] Cowling recasts the platonic formulation in terms of causal "activity"–"Necessarily, some entity x exists if and only if x is causally active" (2013, 307)–while notoriously Molnar do so in terms of causal "operativity" (2000, 77).

[11] For a similar formulation see Martin (1996), Molnar (2000, 78), Jackson & Petitt (1990, 115-116), and Lipton (1993).

[12] Notice that if one holds the view properties are just clusters of powers (see Mumford 1998; 2007; *inter alia*), the current example would not be explicative of this second formulation.

[13] The same definition is inherited in Armstrong (2004), where it is discussed in terms of counterfactual difference (2004, 77). A somehow similar formulation of the principle in terms of difference-making can also be found in Mumford: "for any intrinsic non-abstract property P, P exists if and only if there are circumstances C in which the instantiations of P have causal consequences" (2004, 190).

totality facts in passing the Eleatic test. Nevertheless, we believe that under no interpretation of the test this alleged superiority can be established.

Armstrong is sceptical of the possibility of totality facts being power-bestowing factors.[14] He defends the reality of totality facts via the weaker Eleatic formulation, which only requires that totality facts make a causal difference. In fairness, Armstrong does not discuss his defence at length, but the idea is roughly the following. Let us suppose that by saying that there are only two pots on my desk I denote a (totality) fact [there are two pots on this table and no more]. This particular totality fact makes a causal difference because, had this fact not been obtained, the table would have contained more (or less) pots. The same goes for the "biggest, world-embracing" totality fact that actually obtains:

> had [it] not obtained, the world would have been bigger or smaller. If smaller, then this would presumably have made a difference, if only here and there, to the way the remainder of reality behaved. If larger, presumably that would have made a difference also. (Armstrong 2004, 77)

Similarly, the fact that the current was 100 milliamperes and no more does indeed make a difference to the fact that someone was burned rather than killed. Essentially, however, it is not so evident that totality facts surpass negative facts according to this formulation. Molnar notices (2000, 78-79) that we sometimes say of an absence or a negative object that is causal in the sense captured by Armstrong. Because Molnar thinks of this sense as an inappropriate reading of the Eleatic principle, he concludes that negative facts do not pass the test, and hence are not real. Armstrong thinks his reading is legitimate, and according to this reading totality facts are indeed Eleatically kosher. But, on the same ground, so are negative facts. Suppose we say that the absence of 100ml of water per day causes a plant to die. We are certainly denoting a negative fact, but we might agree that the only powers involved in the possible death of the plant are those involved in the biochemical process—viz. the negative fact is not causally operative. A negative fact is, however, causally relevant in that it can explain why the plant died rather than dried. Assuming that an absence of 50ml of water per day would have dried, rather than killed, the plant, the negative fact can be deemed Eleatically kosher because it makes a causal difference to the world.

---

[14] Armstrong does not offer any argument against using the first formulation to assess the reality of totality facts. He merely briefly claims that: "A difficulty that may be (indeed, should be) proposed is that totality states of affairs fail the Eleatic Stranger's plausible demand (Sophist, 247e) that it is a mark of the real that it should bestow power. That all things of a certain sort are indeed all of that sort does not seem to be a power-bestowing factor in the way the world proceeds" (2004, 76).

Tugby's argument for the reality of totality facts is more optimistic than Armstrong originally intended. Totalities can be deemed real because they can bestow causal powers. So, let us examine the other formulation of the Eleatic principle and grant more generally, alongside Tugby and Armstrong, that negative facts lack causal powers.[15] It is relatively easy to see why many are unconvinced by the possibility of negative facts bestowing causal powers and why some others, on the same ground, might be keen on attributing powers to totality facts. When we say "the absence of clouds in the sky", we denote an absence-negative fact that [there are no clouds in the sky]. Like any facts, whether positive or not, negative facts are not the kind of things that can instantiate powers directly.[16] At best, particulars, not facts, instantiate powers. So, if negative facts bestow powers, it is because their constituents do so. Compositionalism, indeed, has it that facts are complexes that contain properties/powers and (thin) particulars as *parts* (Armstrong 1986). But the constituents of our negative fact, the clouds, do not exist and consequently cannot bestow causal powers. So negative facts do not bestow causal powers.[17]

On the other hand, positive first-order facts have "positive" constituents, i.e. particulars, their properties, and their powers.[18] Assuming powers to be parts of first-order facts and first-order facts to be parts of the totality fact, the totality bestows powers (indirectly) in virtue of constituent facts (indirectly) bestowing powers—by the transitivity of part-of relation.[19] Nevertheless, attributing the parts' powers to the fact they compose is unlovely in the first place. The difficulty lies in the transitivity of part-of. As Frege first adumbrates, if Vesuvius consists of solidified lava, and Vesuvius is part of a fact, "the fact would thus also consist of solidified lava. This would not seem correct to me" (1919, 20). The problem with the fact [Vesuvius' being a volcano] is that, in order to consist of lava, it must be true that facts have some properties, such as extension, that we won't attribute to facts *qua* facts in the first place. Likewise, suppose that a particular glass is fragile, and both the glass and its fragility are part of the fact [this glass' being fragile]. Transitivity then has it that [this glass' being fragile] is itself fragile. But this is absurd. Some, like Vallicella (2000) and Lowe (1998), take the transitivity problem as evidence that

---

[15] On the alleged lack of causal powers by absences, see Hall (2004), and Dowe (2001; 2009), *inter alia*.

[16] This can be disputed, especially if, following Armstrong (1997, 169), we take facts to be particulars.

[17] A similar argument for the conclusion that absences cannot be truthmakers can be found in Dodd (2007, 388).

[18] We use *parts* and *constituents* interchangeably. For a distinction between the former and the latter see Russell (1919, 278) and Armstrong (1989, 92). For a discussion within the context of the debate over Extensionalism and non-Extensionalism about facts' composition see Betti (2015, 68-70).

[19] Arguably, any fundamental parthood relation is transitive. For an extensive defence of transitivity, see Varzi (2006b).

Compositionalism is guilty of misapplying mereological principles to objects whose components belong to a different ontological category altogether.[20] If the mereological principle for facts' composition yields the wrong result, totality facts cannot bestow powers in the way indicated above.

In fairness, some have followed the late Armstrong (1989, 88-93; 2004, 141) in denying Compositionalism, arguing instead that facts are not mereological sums but unified objects—viz. their composition is unmereological or non-mereological—that differ from the collection of their constituents, taken separately.[21] Within this camp, some, like Elder-Vass (2010), claim that in virtue of being non-mereological complexes, the powers of the facts are *distinctive* and "over and above" that of the parts. Setting aside the alleged plausibility of the notion of non-mereological composition—notoriously controversial—we believe that trying to establish that facts are bearers of distinctive powers by using such a notion is problematic for at least two reasons.[22]

Firstly, non-mereological composition is introduced as a primitive notion to explain how given a property F and a (thin) particular *a* we get that [*a* is F]. But then, as Jago has pointed out:

> it can't be intrinsically objectionable to accept a further primitive notion of non-mereological composition which takes F and *a* and gives us the state of affairs that *a* isn't F". (Jago 2018, 206-207)[23]

Hence, if we accept that the non-mereological composition *alone* is responsible for conferring powers to positive facts—on the same ground––friends of negative facts can equally make use of their antithetical notion to establish that also negative facts bestow causal powers.

Secondly, even if positive facts acquire their powers via non-mereological composition in a way that does not grant the opponent the same advantage, it is not at all clear that totality facts would acquire *distinctive* causal powers, i.e. powers over and above the powers bestowed by first-order facts. This is obvious if we consider what totality facts are: complexes of

---

[20] For more recent criticisms of Compositionalism in Armstrong's account of facts see Bynoe (2011).
[21] On the ontological distinctiveness between sums or aggregates, and unities see also Russell (1903, 136).
[22] Non-mereological composition is generally met with skepticism. Lewis famously denies any room for non-mereological composition (1992). Further discussions about Armstrong's compositional pluralism and the issues surrounding non-mereological composition can be found in McDaniel (2019), Maurin (2013), Vallicella (2000), Simons (2009), and Betti (2015).
[23] See also Barker and Jago (2012, 3-4). A similar point is made by Tallant (2017, 56).

first-order facts plus the high-order negative fact that first-order facts are all there are. For a big fact of this sort to have *distinctive* causal powers, something more than the powers of the first-order facts must be available. On the one hand, the high-order negative fact cannot contribute to any power because, as a limit-fact, its negativity is no different from an absence-negative fact. And if the latter does not bestow powers, the former does not either.[24] On the other hand, it is not clear that we can retrieve more powers from the alleged powers of positive first-order facts. Even if first-order positive facts bestow powers, it does not follow that these powers are bestowed by the totality fact rather than by first-order facts. Indeed, what the totality fact itself states is precisely that there are no more facts, and so no more powers, than the one appearing in the positive conjunction.

The same observation can be made in terms of the "totality" relation. As said above, this relation serves to total the first-order facts and so, alongside these facts, serves to ground the truth that there are no other objects. Consider Armstrong's example of the property *being an electron*. This property (i) gives electrons their powers (e.g. to repel like-charged particulars) and (ii) totals the aggregate of electrons (2004, 73). But totalling is not a distinctive power over and above the power conferred by the property of being an electron.[25] Totalling might be relevant for the fact that no non-electron is a member of the aggregate, but it does not confer non-electrons any power such that they are not members of the aggregate; nor does it "do" anything so that nothing other than the aggregate exists. Quite the opposite, the totality relation is in place *only because* there are no more things instantiating the property than there actually are—i.e. it is because non-electrons already do not have electrons-powers that they are not *all* by the property.

It seems that totality facts cannot have *distinctive* casual powers; if they have causal powers, they bestow the powers of first-order facts only. Consequently, only the first-order facts satisfy the Eleatic principle, not the totality. And if Tugby's totality fact does not confer any more causal powers than those needed to reach the current of 100 milliamperes, then Tugby's totality fact is no more ontologically acceptable than a negative fact vis-à-vis bestowing powers.

---

[24] Armstrong seems to acknowledge this point when he claims that "It is important to realise the 'no more' that these facts or states of affairs involve should not be thought of as additions of being. 'No more', after all, is the rejection of any addition" (2006, 245-246).

[25] In other words, totalling is not something that the property does but, as Armstrong puts it "where the aggregate is, the corresponding property has instances, and has them no where else" (2004, 73).

To sum up, solely on the Eleatic considerations, we cannot establish the superiority of totality facts over negative facts. They are either both unacceptable or equally acceptable.

3.   Totality and Explanation

Even though totality and negative facts seem equally acceptable, partial negativity might still be deemed more ontologically acceptable than full negativity. Armstrong (2004, 70) famously argues that this is because partially negative fact removes the need for many fully negative facts, and so further negative truthmakers. Hence, according to the principle of ontological parsimony, totality facts as truthmakers are a more acceptable ontological posit than fully negative facts. We will now argue that that is not the case.

We believe that anyone offering explanations based on totality facts should also endorse the following principle:

> (Closure): For all complexes of positive first-order facts T_i, and some fact v, T_i explains v iff:
> i)   there is no complex of positive first-order facts T_j that includes T_i and that explains v.
> ii)   there is no complex of positive first-order facts T_h that is included in T_i and that explains v.

Why is (Closure) an appropriate testbed for totality facts? Because (Closure) encapsulates two widely endorsed features of explanation—whether or not based on totalities—and, at the same time, one crucial task of truthmaking strategies—whether or not totalities are taken as truthmakers.

(Closure) captures the (i) non-monotonicity and the (ii) minimality of explanation (see Woodard 2003; Fine 2012a), expressed here in terms of totalities, i.e. complexes of first-order facts. Firstly, explanations are non-monotonic in that information contained in the explanans must be relevant to explaining the explanandum. Hence, the fact that [the mug is maroon] obtains explains why the fact that [the mug is red] obtains, but it is not the case that [the mug is maroon and that it is made of metal] explains why [the mug is red] obtains. The mug being made of metal does not contribute to the mug being red. Explanatory relations, like determination relations, are non-monotonic in this sense. Secondly, explanations are minimal in that every piece of information in the explanans must jointly suffice to

bring about the fact explained. The fact that an object is round and elastic explains why it is a football, but it is not the case that the fact that it is only elastic explains why it is a football. The fact that it is round cannot be subtracted from the explanans without it ceasing to explain the explanandum. Hence, jointly expressed in (Closure), minimality and non-monotonicity work to ensure that no more and no less than what is needed to explain the explanandum is contained in the explanans.

A totality truthmaking strategy, such as the one marshalled by Tugby or Armstrong, should adhere to (Closure) not only on the grounds of non-monotonicity and minimality of explanation but also because, crucially, any truthmaking strategy should result in the individuation of truthmakers that are *exact*, i.e. they must guarantee *v's* truth as well as being wholly or strictly relevant to it (Fine & Jago 2019). For example, take the fact that x is an apple and assume that x is an apple is made true by facts about its microstructure y. Let us assume a further physical fact that makes true facts about the apple's microstructure. For example, we can notice that x has a microstructure y is true iff y has definite values O_i for any quantity O measured. And, further, that y has a definite value O_i for any quantity O iff there is a state $|\psi\rangle$ that is an eigenstate of O—the eigenvalue-eigenstate link principle. While facts about the microstructure y are strictly relevant to x being an apple, facts about eigenvalue-eigenstate are not, because they make true many other facts about many other macroscopic entities. That is, the eigenvalue-eigenstate link principle is non-discerning between the truth about the apple and other truths—viz. about this chair, that table, and so on. However, to discern truths in this way is the job of *any* truthmaking strategy (Armstrong 2004, 18; Tahko 2013, 336-7; *inter alia*), and insofar as the strategy of the friends of totality facts is a truthmaking strategy, they shall meet the task too. Put in another way, (closed) totality facts are *exact* truthmakers (Fine 2017; Fine & Jago 2019; *inter alia*) if they are truthmakers at all: they must guarantee *v's* truth as well as being wholly relevant to it.[26] Thus, as Tugby suggests, we should expect them to satisfy (Closure). Unfortunately, totality facts, such as the one provided in Tugby's electric current example, violate both non-monotonicity and minimality.

---

[26] The case for exact truthmaker semantics can be traced back to van Fraassen (1969), but has been developed more extensively by Fine (2012b, 2014, 2016) and, further, by Correia (2016), Yablo (2014, 2018), van Rooji (2017), *inter alia*. The benefits of the exact truthmaking relation have been advocated both in the realm of everyday semantic theorising (see Moltmann 2017; *inter alia*) and in its foundational ground, as a *ur*-theory capable of constructing possible worlds semantics, standard situation semantics, and the familiar notion of inexact truthmaking (Fine 2020). For an argument that inexact, rather than exact, truthmaking semantics should play a foundation role see Deigan (2020).

Let us take T_i to be the mereological complex of positive first-order facts consisting of (units of electrical current, burning). As per the standard reading, the mereological object consisting of the units of electrical current *total* the property of burning, so that 100 units and no more explains why I was burned rather than killed. Given minimality, there should be no T_h that is included in T_i that would explain this occurrence. Put simply, nothing short of 100 milliAmps should explain why I was burned rather than killed. Crucially, however, there *is* a T_h included in T_i such that it violates minimality. The mereological object consisting of 99 milliAmps, like the previous one, explains why I was burned rather than killed. Given that both T_h and T_i explain why I survived, an explanation based on T_i fall short of minimality. A case can be made for the conclusion that T_i falls short of non-monotonicity too. Basic knowledge from electronics has it that electrical current cannot flow without voltage, whether deadly or not. A live wire is a wire where voltage is present and current flows. So given T_i (units of electrical current, burning), we can take T_j to be the mereological complex of positive first-order facts including T_i and consisting of (units of electrical current, burning, *and* units of voltage). Put simply, we now have a mereological object that includes the former and that total the property of burning. It makes sense to generate something like T_i because there is no current without voltage, facts about the first are strongly connected with the second. However, this cannot be right: it is not the voltage but the current and Amp value that explain why I was burned rather than killed. The fact that there is voltage—that the wire is live—is evidence supporting the explanation of me being burned because of the current, but it is not itself part of the explanation (we can, in fact, have voltage without current).

There is something significant we can learn from the failure of the totality fact. Going back to the current example one last time, Tugby takes the totality fact to be the total current fact, and we made this more explicit by equating that fact to a complex of positive first-order facts consisting of (units of electrical current, burning). The obvious upshot is that we have neglected that the totality fact is a complex of first-order positive facts *plus* a high-order negative fact that *no more is there than what there is*. The complex T_i should have been what *total* the property of burning (the units of current) together with the high-order fact that nothing else totals the property. Reintroducing the negative fact has different consequences depending on how we interpret its role. If it is interpreted as representing the fact that the -*ally* relation instantiated by the first-order facts is such that these are the *sole* facts that total the property (see Armstrong 1989, 93-94), then nothing more *and* nothing less total the property. Thus, interestingly, by adding the negative component back into the totality fact,

minimality and non-monotonicity are easily met; they are packed, so to speak, in the negative component. If nothing else totals the property of burning, nothing more and nothing less than the current magnitude of the property explain why I was burned rather than killed because those are the sole facts that can do so.

Now, however, we are saying that the only way to establish the explanatory power and the truthmaking role of totality fact is to appeal specifically to its negative part. In virtue of the high-order negative component, seemingly, totality facts regain their good standing. But this cannot be right if, per our original hypothesis, partially negative facts are better explanans than fully negative facts. This is because the negative component of a totality fact *is* a fully negative fact. Given that no other component of the totality contributes to it being a good explanans, we should then conclude that partially negative facts are thus no explanatorily better than fully negative facts.

However, we can entertain a more liberal interpretation, such that the limit-negative fact expresses the fact that nothing more *but not* nothing less than the current magnitude of the property explains why I was burned rather than killed. It is easy to see how this proposal offends against parsimony. Suppose our office contains only a chair, a desk, and a lamp. The corresponding totality fact is the aggregate of first-order facts—e.g. there being a chair in the office, there being a lamp in the office, and so on—plus the high-order negative facts that these are all the facts.[27] Suppose further that it is true that "There is no red carpet in this office" and that the totality fact is its truthmaker. Plausibly, the totality is a truthmaker for "There is no red carpet in this office" only if "There is no red carpet in this office" is true *in virtue of* the totality existing and guaranteeing its truth. The totality fact can guarantee the truth of the proposition only if there are no first-order facts such that it would make the proposition false—e.g. it necessitates the truth of "There is no red carpet in this office" only if [there is a red carpet] is not part of the aggregate. However, while the high-order negative fact so interpreted can establish that these are all the facts—it sets a limit—it does not, however, guarantee that the totality does not also contain such false makers. Now, these false makers can either be ruled out "from the outside" or "from the inside". If they are ruled out from the outside, that must be because "There is no red carpet in this office" is true. But then the direction of explanation is reversed: it is the negative existential that explains why these are all the facts, rather than the other way around.

---

[27] Alternatively, we can represent the totality as: *T* (aggregate of furniture, being our office furniture).

However, to rule out such false-makers from the inside, and so to guarantee the truth of "There is no red carpet in this office"—or of "there is no cabinet in this office", or of any other negative claim true of our office—requires assuming in advance that the totality contains the negative facts that the aggregate does not contain [there being a red carpet]—or a cabinet, a giraffe, and so on. But this is clearly an unwelcome result. If that is the case, the truthmaking role of totality facts is not supplied anymore by the one high-order negative fact, but by the many negative facts regarding what the positive totality does not contain. Thus, the alleged economic advantage vanishes because, after all, *pace* Armstrong, we still need many truthmakers and not one. On grounds of (quantitative) parsimony, friends of negative facts have at least equal standing.

4.   Conclusions

In this paper, we have argued for a modest conclusion: that totality facts are no more ontologically respectable than negative facts. Although the conclusion is not novel (see Molnar 2000), we have proceeded in a novel way: by granting, instead of rejecting, the distinction between partial and full negativity. We hope to have shown that, nevertheless, there are reasons to reject the alleged ontological superiority of the former over the latter and, hence, of totality facts over traditional negative facts. We have proceeded in a systematic fashion by, first, arguing on the base of Eleatic considerations that totality facts and fully negative facts are either equally ontologically unacceptable or equally ontologically legitimate. Secondly, we have presented a truthmaking argument for the conclusion that adopting totality facts does not yield the desired alleged ontological parsimony. Our conclusion is that facts that are partially negative seem no better than facts that are fully negative.

Acknowledgments

REFERENCES

Armstrong, David M. 1986. *The Nature of Possibility*. *Canadian Journal of Philosophy* 16 (4): 575-594.

———. 1989. *A Combinatorial Theory of Possibility.* Cambridge University Press.

———. 1997. *A World of States of Affairs.* Cambridge: Cambridge University Press.

———. 2007. "Truthmakers for Negative Truths, and for Truths of Mere Possibility." In *Metaphysics and Truthmakers*, edited by J. M. Monnoyer, 99. Ontos Verlag.

———. 2004. *Truth and Truthmakers.* Cambridge: Cambridge University Press.

———. 2006. "Reply to Heil." *Australasian Journal of Philosophy* 84 (2): 245-247.

Barker, Stephen, and Mark Jago. 2012. "Being Positive About Negative Facts." *Philosophy and Phenomenological Research* 85 (1): 117-138.

Beall, J. C. 2000. "On Truthmakers for Negative Truths." *Australasian Journal of Philosophy* 78 (2): 264-268.

Betti, Arianna. 2015. *Against Facts.* Cambridge, MA, USA: The MIT Press.

Bynoe, William. 2011. "Against the Compositional View of Facts." *Australasian Journal of Philosophy* 89 (1): 91-100.

Cameron, Ross P. 2008. "How to be a Truthmaker Maximalist." *Noûs* 42: 410-421.

Correia, Fabrice. 2016. "On the Logic of Factual Equivalence." *The Review of Symbolic Logic* 9 (1): 103-122.

Cowling, Sam. 2013. "Ideological Parsimony." *Synthese* 190 (17): 3889-3908.

Deigan, Matthew. 2020. "A Plea for Inexact Truthmaking." *Linguistics and Philosophy* 43 (5): 515-536.

Dodd, Julian. 2007. "Negative Truths and Truthmaker Principles." *Synthese* 156 (2): 383-401.

Dowe, Phil. 2001. "A Counterfactual Theory of Prevention and 'Causation' by Omission." *Australasian Journal of Philosophy* 79 (2): 216-226.

———. 2009. "Absences, Possible Causation, and the Problem of Non-Locality." *The Monist* 92 (1): 23-40.

Elder-Vass, Dave. 2010. *The Causal Power of Social Structures: Emergence, Structure and Agency*. Cambridge University Press.

Fine, Kit. 2012a. "Guide to Ground." In *Metaphysical Grounding*, edited by F. Correia and B. Schnieder, 37-80. Cambridge University Press.

———. 2012b. "Counterfactuals Without Possible Worlds." *The Journal of Philosophy* 109 (3): 221-246.

———. 2014. "Permission and Possible Worlds." *Dialectica* 68 (3): 317-336.

———. 2016. "Angellic Content." *Journal of Philosophical Logic* 45 (2): 199-226.

———. 2020. "Yablo on Subject-Matter." *Philosophical Studies* 177 (1): 129-171.

Fine, Kit, and Mark Jago. 2019. "Logic for Exact Entailment." *Review of Symbolic Logic* 12 (3): 536-556.

Frege, Gottlob. 1919. "Letter to Wittgenstein 28.6.1919."*Grazer Philosophische Studien* 33/34: 19-20.

Hall, Ned. 2004. "Two Concepts of Causation." In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul, 225-276. MIT Press.

Heil, John. 2000. "Truth Making and Entailment." *Logique and Analyse* 169-170: 231-242.

Ingram, David. 2016. "Platonism, Alienation, and Negativity." *Erkenntnis* 81 (6): 1273-1285.

Jackson, Frank, and Philip Pettit. 1990. "Program Explanation." *Analysis* 50 (2): 107-117.

Jago, Mark. 2018. "From Nature to Grounding." In *Reality and its Structure*, edited by Ricki Bliss and Graham Priest, 199-216. New York: Oxford University Press.

Kukso, Brian. 2006. "The Reality of Absences." *Australasian Journal of Philosophy* 84 (1): 21-37.

Lipton, Peter. 1993. "Contrastive Explanation." In *Explanation*, edited by David H. Ruben, 207-227. Oxford University Press.

Lewis, David K. 1992. "Critical Notice of David Armstrong, A Combinatorial Theory of Possibility." *Australasian Journal of Philosophy* 70: 211-224.

Lowe, E. J. 1998. *The Possibility of Metaphysics: Substance, Identity, and Time*. Clarendon Press.

Martin, C. B. 1996. "How It Is: Entities, Absences and Voids." *Australasian Journal of Philosophy* 74: 57-65.

Maurin, Anna-Sofia. 2013. "States of Affairs and the Relation Regress." In *The Problem of Universals in Contemporary Philosophy*, edited by Gabriele Galluzzo, Michael J. Loux, and Thomas M. Crisp, 111-128. Oxford University Press.

McDaniel, Kris. 2009. "Structure-making." *Australasian Journal of Philosophy* 87 (2): 251-274.

Merricks, Trenton. 2007. *Truth and Ontology*. Oxford: Clarendon Press.

Molnar, George. 2000. "Truthmakers for Negative Truths." *Australasian Journal of Philosophy* 78: 72-86.

Moltmann, Friederike. 2017. "Levels of Linguistic Acts and the Semantics of Saying and Quoting." In *Interpreting Austin: Critical Essays*, edited by Savas L. Tsohatzidis, 34-59. Cambridge University Press.

Mumford, Stephen. 1998. *Dispositions*. Oxford: Clarendon Press.

Rodriguez-Pereyra, Gonzalo. 2005. "Why Truthmakers." In *Truthmakers: The Contemporary Debate*, edited by Helen Beebee and Julian Dodd, 17-31. Oxford: Clarendon Press.

Russell, Bertrand. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

———. 1918. "The Philosophy of Logical Atomism." *Monist* 28: 495-527.

———. 1919. "On Propositions: What They Are and How They Mean." *Proceedings of the Aristotelian Society, Supplementary Volume* 2: 1-43.

———. 1989. *Logic and Knowledge.* Edited by Robert Charles Marsh. London: Unwin Hyman.

Schaffer, Jonathan. 2010. "The Least Discerning and Most Promiscuous Truthmaker." *The Philosophical Quarterly* 60 (239): 307-324.

Simons, Peter. 2009. "Why There Are No States of Affairs." In *States of Affairs*, edited by Mauro Antonelli Reicher, 111-128. Ontos.

Tahko, Tuomas E. 2013. "Truth-Grounding and Transitivity." *Thought: A Journal of Philosophy* 2 (4): 332-340.

Tallant, Jonathan. 2010. "There's No Existent like 'No Existence' like No Existent I Know." *Philosophical Studies* 148 (3): 387-400.

———. 2017. *Truth and the World: An Explanationist Theory.* London: Routledge.

Tugby, Matthew. 2017. "Modal Platonism and the Problem of Negativity." *Erkenntnis* 44.

Vallicella, William F. 2000. "Three Conceptions of States of Affairs." *Noûs* 34 (2): 237-259.

Van Fraassen, Bas. 1969. "Facts and Tautological Entailments." *The Journal of Philosophy* 66 (15): 477-487.

Van Rooij, Robert. 2017. "A Fine-Grained Global Analysis of Implicatures." In *Linguistic and Psycholinguistic Approaches on Implicatures and Presuppositions*, edited by Salvatore Pistoia-Reda and Filippo Domaneschi, Palgrave Macmillan.

Varzi, Achille. 2006a. "The Talk I Was Supposed to Give…" In *Modes of Existence: Papers in Ontology and Philosophical Logic*, edited by Andrea Bottani and Richard Davies, 131-152. Ontos Verlag.

——— 2006b. "A Note on the Transitivity of Parthood." *Applied Ontology 1* (2): 141-146.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press.

Yablo, Stephen. 2014. *Aboutness*. Princeton University Press.

———. 2018. "Reply to Fine on Aboutness." *Philosophical Studies* 175: 1495-1512.

ABSTRACTS (SAŽECI)

Are there "Moral" Judgments?

David Sackris
Arapahoe Community College

Rasmus Rosenberg Larsen
University of Toronto, Mississauga

ABSTRACT

Recent contributions in moral philosophy have raised questions concerning the prevalent assumption that moral judgments are typologically discrete, and thereby distinct from ordinary and/or other types of judgments. This paper adds to this discourse, surveying how attempts at defining what makes moral judgments distinct have serious shortcomings, and it is argued that any typological definition is likely to fail due to certain questionable assumptions about the nature of judgment itself. The paper concludes by raising questions for future investigations into the nature of moral judgment.

Keywords: metaethics; moral judgment; judgment; ontology.

Postoje li „moralni" sudovi?

David Sackris
Arapahoe Community College

Rasmus Rosenberg Larsen
University of Toronto, Mississauga

SAŽETAK

Nedavne rasprave u moralnoj filozofiji dovode postavljaju pitanja oko prevladavajuće pretpostavke da su moralni sudovi tipološki diskretni i da se stoga razlikuju od običnih i/ili drugih vrsta sudova. Ovaj rad se nadovezuje na ovu raspravu, razmatrajući nedostatke različitihi pokušaja definiranja onoga što čini neke sudove moralnima, te se tvrdi da svaka tipološka definicija vjerojatno neće uspjeti zbog upitnih pretpostavki o prirodi sudova. Rad zaključuje postavljanjem pitanja koja se odnose na buduća istraživanja prirode moralnih sudova.

Ključne riječi: metaetika; moralni sudovi; sud; ontologija.

Free Will as An Epistemically Innocent False Belief

Fabio Tollon

Bielefeld University, Stellenbosch University, Centre for Artificial Intelligence Research (CAIR), Pretoria, South Africa

ABSTRACT

In this paper I aim to establish that our belief in free will is epistemically innocent. Many contemporary accounts that deal with the potential "illusion" of freedom seek to describe the pragmatic benefits of belief in free will, such as how it facilitates or grounds our notions of moral responsibility or basic desert. While these proposals have their place (and use), I will not explicitly engage with them. I aim to establish that our false belief in free will is an epistemically innocent belief. I will endeavour to show that if we carefully consider the circumstances in which particular beliefs (such as our belief in free will) are adopted, we can come to better appreciate not just their psychological but also their epistemic benefits. The implications, therefore, for future investigations into the philosophy of free will are that we should consider whether we have been too narrow in our pragmatic defences of free will, and that we should also be sensitive to epistemic considerations.

Keywords: Free will; epistemic innocence.

Slobodna volja kao epistemički nevino lažno vjerovanje

Fabio Tollon

Bielefeld University, Stellenbosch University, Centre for Artificial Intelligence Research (CAIR),
Pretoria, South Africa

SAŽETAK

U ovom radu želim utvrditi da je naše vjerovanje u slobodnu volju epistemički nevino. Mnogi suvremeni prikazi koji se bave potencijalnom "iluzijom" slobode nastoje opisati pragmatične dobrobiti vjerovanja u slobodnu volju poput toga kako ona olakšava ili utemeljuje naše predodžbe o moralnoj odgovornosti ili osnovnoj zasluzi. Iako ovi prijedlozi imaju svoje mjesto (i korist) neću se eksplicitno baviti njima. Želim utvrditi da je naše lažno vjerovanje u slobodnu volju epistemički nevino vjerovanje. Nastojat ću pokazati da ako pažljivo razmotrimo okolnosti u kojima su određena uvjerenja (kao što je naše vjerovanje u slobodnu volju) usvojena, možemo bolje cijeniti ne samo njihove psihološke, već i njihove epistemičke koristi. Stoga su implikacije za buduća istraživanja filozofije slobodne volje te da bismo trebali razmotriti jesmo li bili preusko definirali pragmatične obrane slobodne volje te bismo, također, trebali biti osjetljivi na epistemička razmatranja.

Ključne riječi: Slobodna volja; epistemička nevinost.

Why Parent Together?

Marcus William Hunt
Concordia University Chicago

ABSTRACT

The paper offers an account of co-parenthood according to which co-parents are parent and child to one another. The paper begins by reviewing extant theories of the value of being a parent, to see whether the value of co-parenthood is reducible to this. Finding that it is not, I briefly elaborate a theory of parenthood on which parents are those who create persons. Using Aristotle's four causes as a helpful prism, I outline how parents are the cause of their child, and how in causing a child together co-parents become parent and child to one another. For instance, since parents create

children by offering themselves as models to be copied, co-parents should enjoy the best type of friendship with one another, each treating the other's flourishing as a human person as their end. I suggest that co-parenthood contains parenthood virtually, that the co-parents' love of their child is a manifestation of their love for one another, that the teleological fulfilled state of the friendship between parent and child exists in the friendship of co-parents.

Keywords: Aristotle; co-parent; family ethics; parent; solo-parent.

Zašto biti su-roditelj?

Marcus William Hunt
Concordia University Chicago

SAŽETAK

U radu se nudi prikaz su-roditeljstva prema kojem su su-roditelji jedno drugom i roditelj i dijete. Rad započinje pregledom postojećih teorija o vrijednosti roditeljstva kako bi se vidjelo može li se vrijednost su-roditeljstva svesti na to. Utvrdivši da se ne može, ukratko elaboriram teoriju roditeljstva po kojoj su roditelji ti koji stvaraju osobe. Koristeći Aristotelova četiri uzroka kao korisnu prizmu, ocrtavam kako su roditelji uzrok svog djeteta i kako u stvaranju djeteta zajedno su-roditelji postaju roditelji i dijete jedno drugome. Na primjer, s obzirom na to da roditelji stvaraju djecu kao modele koje treba kopirati, su-roditelji bi trebali uživati u najboljoj vrsti prijateljstva jedni s drugima, tretirajući procvat onog drugog kao ljudske osobe kao svoj cilj. Predlažem da su-roditeljstvo sadrži roditeljstvo virtualno, da je ljubav su-roditelja prema djetetu manifestacija njihove ljubavi jednih prema drugima te da teleološki ispunjeno stanje prijateljstva između roditelja i djeteta postoji u prijateljstvu su-roditelja.

Ključne riječi: Aristotel; su-roditelj; obiteljska etika; roditelj; solo-roditelj.

Morality without Categoricity

Elizabeth Ventham
University of Salzburg, Austria

ABSTRACT

This paper argues that an agent's moral obligations are necessarily connected to her desires. In doing so I will demonstrate that such a view is less revisionary—and more in line with our common-sense views on morality—than philosophers have previously taken it to be. You can hold a desire-based view of moral normativity, I argue, without being (e.g.) a moral relativist or error theorist about morality. I'll make this argument by showing how two important features of an objective morality are compatible with such a desire-based account: 1) morality's authoritative nature, 2) our ability to condemn immoral agents.

Keywords: Meta-ethics; practical reasoning; hypothetical imperatives; desire; moral realism.

Moralnost bez kategoričnosti

Elizabeth Ventham
University of Salzburg, Austria

SAŽETAK

U članku se tvrdi da su moralne obveze djelatnice nužno povezane s njezinim željama. Pri tome pokazujem da je takva perspektiva manje revizionistička – i više u skladu s našim zdravorazumskim stajalištima o moralu – nego što su ga filozofi prethodno smatrali. Tvrdim da možemo imati gledište na normativnost koje je utemeljeno na željama, a da nismo (na primjer) moralni relativisti ili teoretičari pogreške o moralu. Iznosim ovaj argument pokazujući kako su dvije važne značajke objektivnog morala kompatibilne s takvim gledištem temeljenim na želji: 1) autoritativna priroda morala, 2) naša sposobnost da osudimo nemoralne djelatnike.

Ključne riječi: Meta-etika; praktično zaključivanje; hipotetički imperativ; želje; moralni realizam.

Something Negative about Totality Facts

Andrea Raimondi

Thapar Institute of Engineering and Technology, Patiala, India

ABSTRACT

Armstrong famously argued in favour of introducing totality facts in our ontology. Contrary to fully negative (absence) facts, totality facts yield a theory of "moderate" or "partial" negativity, which allegedly provides an elegant solution to the truthmaking problem of negative claims and, at the same time, avoids postulating (many) first-order absences. Friends of totality facts argue that partial negativity is (i) tolerable vis-à-vis the Eleatic principle qua mark of the real, and (ii) achieves a significant advantage in terms of ontological parsimony. But are totality facts, which are partially negative, really more ontologically acceptable than fully negative facts? In this paper, we argue that, comparatively, the case for totality facts is weaker than commonly assumed and that, ultimately, the answer is negative.

Keywords: Totality facts; Armstrong; negative facts; truthmaking; causal powers.

Nešto negativno o totalnim činjenicama

Andrea Raimondi

Thapar Institute of Engineering and Technology, Patiala, India

SAŽETAK

Armstrong je poznat po zagovaranju uvođenja totalnih činjenica u našu ontologiju. Suprotno potpuno negativnim činjenicama (odsustvu), totalne činjenice pružaju teoriju ili djelomične negativnosti, koja navodno pruža rješenje za utemeljenja istinitosti negativnih tvrdnji te istovremeno izbjegava postuliranje (mnogih) negativnih (odsustva) tvrdnji prvog reda. Zagovornici totalnih činjenica tvrde da je djelomična negativnost (i) prihvatljiva u odnosu na elejsko načelo kao oznake stvarnog, i (ii) da postiže značajnu prednost u smislu ontološke ekonomičnosti. Međutim, jesu li totalne činjenice, koje su djelomično negativne, zaista ontološki prihvatljivije u odnosu na potpuno negativne činjenice? U ovom radu, usporedno gledano, tvrdimo da je argument za totalne činjenice slabiji nego što se uobičajeno pretpostavlja i da je konačni odgovor negativan.

Ključne riječi: Totalne činjenice; Armstrong; negativne činjenice; utemeljenje istinitosti; uzročne moći.

Translated by Iva Martinić (Rijeka)

Proofread by Iva Martinić (Rijeka)

(AB)7

AUTHOR GUIDELINES

Publication ethics

EuJAP subscribes to the publication principles and ethical guidelines of the Committee on Publication Ethics (COPE).

Submitted manuscripts ought to:

- be unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere;

- be approved by all co-Authors;

- contain citations and references to avoid plagiarism, self-plagiarism, and illegitimate duplication of texts, figures, etc. Moreover, Authors should obtain permission to use any third party images, figures and the like from the respective copyright holders. The pre-reviewing process includes screening for plagiarism and self-plagiarism by means of internet browsing and software Turnitin;

- include the DOI links to items in the reference list if available (see the DOI display guidelines at Crossref)

- be sent exclusively electronically to the Editors (eujap@ffri.uniri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format;

- be prepared for blind refereeing: authors' names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in document properties should be removed. Generally, there are two options for anonymizing references to ones own work: 1) Author X (2020) argues that (…); 2) In my previous work (Masked reference), I have argued (…). EuJAP encourages the authors who plan to submit their manuscript to use the former, third person, type of self-citation because it allows the reviewers to check the relevance, adequacy, accuracy, etc. of the references. However, if for some reason the prospective author(s) believe it would behoove them to use the second type of self-citation, they are allowed to do so provided that they offer a short explanation of this decision in the accompanying title page;

- be accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 300 words),

keywords, academic affiliation and full address for correspondence including e-mail address, and, if needed, a disclosure of the Authors' potential conflict of interest that might affect the conclusions, interpretation, and evaluation of the relevant work under consideration;

- be in American or British English;

- be no longer than 9000 words, including references (for Original and Review Articles).

- be between 2000 and 5000 words, including footnotes and references (for Discussions and Critical notices)

We ask authors not to send us more than one manuscript at a time. If a single author would like to submit more than one paper to EuJAP, then, ceteris paribus, they should be submitted at least two months apart.

Norms for publishing with AI

The Journal does not exclude the use of AI generated text. However, all authors (including reviewers and editors) take full responsibility for its factual accuracy and the proper acknowledgement of sources. In the acknowledgement section of your manuscript or the title page (depending on the submission/publication stage) or in other kind of reports you must identify the AI that was used, and the extent of the contribution. For instance, ChatGPT (version or the date when the AI was used).

The contribution level of the AI can be defined as follows:

- negligible – means the AI only made minor changes to the manuscript's style or grammar (this includes using AI for copyediting and similar services);

- modest – means the AI made important suggestions but was not the primary driver of the research or had an essential role in writing the manuscript;

- substantial – means the AI made several crucial suggestions that shaped the research and the manuscript could not have been completed without it.

If the contribution of the AI is "negligible", there is no requirement to mention its usage during the submission or review and publication processes. However, for any other level of contribution, it is expected that authors will report the extent of AI usage. In cases where the AI

contribution is "substantial", authors, reviewers, and editors should provide a comprehensive description of the AI usage and its contributions in a narrative format.

Malpractice statement

If the manuscript does not match the scope and aims of EuJAP, the Editors reserve the right to reject the manuscript without sending it out to external reviewers. Moreover, the Editors reserve the right to reject submissions that do not satisfy any of the previous conditions.

If, due to the authors' failure to inform the Editors, already published material will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue and remove the publication from EuJAP web site and the repository HRČAK.

In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors.

For additional comments, please visit our web site and read our Publication ethics statement (https://eujap.uniri.hr/publication-ethics/). To get a sense of the review process and how the referee report ought to look like, the prospective Authors are directed to visit the *For Reviewers* page on our web site (https://eujap.uniri.hr/instructions-for-reviewers/).

For other policy and related issues, see our website: https://eujap.uniri.hr/author-guidelines/

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style

- contain footnotes and no endnotes

- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations.* Cambridge: Harvard University Press.

Book with multiple authors

T: (Hirstein, Sifferd, and Fagan 2018, 100)

R: Hirstein, William, Katrina Sifferd, and Tyler Fagan. 2018. *Responsible Brains: Neuroscience, Law, and Human Culpability*. Cambridge, Massachusetts: The MIT Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, Richard. 2006. 'The Epistemic Role of Testimony: Internalist and Externalist Perspectives'. In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 77–91. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199276011.003.0004.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, Jennifer, and Ernest Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. "Normative requirements." *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)
R: Skorupski, John. 2010. "Sentimentalism: Its Scope and Limits." Ethical Theory and Moral Practice 13 (2): 125–36. https://doi.org/10.1007/s10677-009-9210-6.

Article with multiple authors in a journal
T: (Churchland and Sejnowski 1990)
R: Churchland, Patricia S., and Terrence J. Sejnowski. 1990. "Neural Representation and Neural Computation." *Philosophical Perspectives 4*. https://doi.org/10.2307/2214198

T: (Dardashti, Thébault, and Eric Winsberg 2017)
R: "Dardashti, Radin, Karim P. Y. Thébault, and Eric Winsberg. 2017. Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity." *The British Journal for the Philosophy of Science* 68 (1): 55–89. https://doi.org/10.1093/bjps/axv010

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013.
http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing 'forthcoming' instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Archiving rights

The papers published in EuJAP can be deposited and self-archived in the institutional and thematic repositories providing the link with DOI to the journal's web pages and/or <u>HRCAK</u>.

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:

International:

individuals € 50

institutions € 100

Croatia:

individuals € 30

institutions € 60

European Journal of Analytic Philosophy is published twice per year.