Nietzschean Moral Error Theory

Patrick Hassan, Cardiff University
hassanp1@cardiff.ac.uk
Final version forthcoming in *History of Philosophy Quarterly*

Introduction

According to moral error theorists, moral judgements attribute particular properties to certain states of affairs, and in doing so purport to accurately describe a feature of the world. However, according to moral error theorists, these properties do not exist, hence, all moral judgements are systematically false. A moral error theory thus has two essential components:

(1) *Semantic Thesis*: moral judgements express beliefs which ascribe moral properties to their subjects.

(2) *Ontological Thesis*: these moral properties to which moral judgements refer do not exist.

Philosophers that have endorsed moral error theory have done so for different reasons. For instance, on the grounds that moral properties are metaphysically queer in some respect(s) (Mackie, 1977; Olson, 2014; Streumer, 2017);[1] that moral properties essentially, but implausibly, presuppose a commitment to categorical 'external reasons' (Williams, 1980; Joyce, 2001); and that evolutionary and/or socio-historical debunking explanations for human behaviour and moral disagreement undermine the need to postulate the existence of 'moral properties' at all (Mackie, 1977; Joyce, 2001, 2006).

Nietzsche has often been associated with moral error theory. Some have interpreted him as endorsing something *close* to it; as a "vital ancestor" (Joyce, 2001: 179; cf. Olson, 2014: 16-17). Others have attributed to Nietzsche a full-blown error theory about all evaluative judgements (Pigden, 2007; Hussain, 2007, 2012; Blackman, 2020); an error theory only about "existing moral discourse" (Sinhababu, 2007: 264, fn. 1; 2015: 280-281; cf. Robertson, 2020: ch. 2, 3 & 11); or an error theory only in specific periods of his writing (Clark & Dudrick, 2007). As we shall see, there are a host of passages

[1] This 'queerness' has been attributed to various (alleged) intrinsic features of moral discourse: (i) the supervenience of moral properties on descriptive properties; (ii) a commitment to motivational internalism; (iii) intuitive knowledge of moral properties; (iv) a commitment to irreducibly normative properties.

from Nietzsche's corpus which offer *prima facie* reasons for an error theoretic interpretation. The *extent* of the appropriateness of this interpretation, however, is a matter of further dispute.

This paper aims to re-consider the evidence for the view that Nietzsche is a moral error theorist, and in doing so engage with Nietzsche's claims about the status of moral judgments through a contemporary lens, to the extent that this is possible. While acknowledging potential concerns over anachronistic exegesis, this paper makes the case that Nietzsche defends a *local* error theory about a particular form of 'morality', but a *global* error theory about value judgements in general is not established by the textual evidence. I defend this view by considering Nietzsche's affinities with Hume, and how they are better harnessed in service of an error-theoretic reading as opposed to alternatives. Doing so will require steering a course between competing existing positions in the secondary literature, explaining how Nietzsche can continue to make evaluative judgements since, like Hume, he draws a distinction between *conventional* evaluative practice—characteristic of herd morality—on the one hand, and a *revisionary* evaluative practice available to an elite few 'higher types' or 'free spirits' on the other. An attempt to *defend* the resulting position is not made here, but the paper does intend to draw attention to how some of the philosophical manoeuvres made can be illuminating for contemporary meta-ethicists beyond Nietzsche scholarship.

1. The Shape of Nietzsche's Moral Anti-Realism

1.1. Anti-Realism about Moral Properties

Moral anti-realism is the denial of the thesis that moral properties exist mind-independently, where *X*'s mind-independence—or 'objectivity'—is understood as *X*'s existence and nature irrespective of beliefs or attitudes about *X*.[2] There is a host of textual evidence from across Nietzsche's corpus which strongly suggest he is an anti-realist about moral properties. To give but a small but representative sample, in *Daybreak* Nietzsche describes our moral judgements as "only images and fantasies based on a

[2] Anti-realism has various forms: it may involve the claim that (a) moral properties do not exist *at all*; or that (b) moral properties exist, but are restricted to a *mind-dependent* reality. This distinction will become crucial later.

physiological process unknown to us" (*D*, §119). Perhaps most clearly indicative of Nietzsche's moral anti realism is the following:

> One knows my demand of philosophers that they place themselves *beyond* good and evil — that they have the illusion of moral judgement *beneath* them. This demand follows from an insight first formulated by me: that *there are no moral facts whatsoever*. Moral judgment has this in common with religious judgement that it believes in realities which do not exist. Morality is only an interpretation of certain phenomena, more precisely a *mis*interpretation. Moral judgement belongs, as does religious judgement, to a level of ignorance at which […] 'truth' […] denotes nothing but things which we today call 'imaginings' (*TI*, 'Improvers': §1)

Nietzsche offers a range of arguments to establish this (at present, broad) anti-realism about moral properties. I shall mention just two here.[3] One argument Nietzsche frequently makes is that intrinsic to the practice of moral judgement are dubious descriptive presumptions (e.g. about agency, the self, or the equality of persons). Such presumptions, he claims, turn out to be mistaken upon critical analysis, making moral judgements incoherent. Nietzsche repeats this objection with respect to free will and moral responsibility increasingly into his writing (e.g. see *D*, §124, §128; *BGE*, §19, §21; *TI*, 'Four Great Errors': §1). Moral judgements essentially presuppose agents to be ultimately responsible for their actions. But such responsibility is a "primeval delusion" (*D*, §116), making such judgements incoherent.

A second argument for anti-realism Nietzsche makes is abductive in nature. He seeks to show that the *best explanation* for moral judgements is not that there are objective moral properties 'out there in the world' which we discover or intuit, but rather that moral judgements have their origins in, and are ultimately reducible to, affective attitudes. This argument has its antecedents in Hume (whom I shall say more about shortly) and especially Herder.[4] In the same way we can determine a type of tree by the fruit it bears, Nietzsche holds that one's moral judgements are the product of the asserter's psychology (e.g. see *D,* §119, §542*; GS,* Preface: §2*; BGE,* §6; *TI*, 'Expeditions': §37): they are "merely a sign language of the affects" (*BGE*, §187) which philosophers in particular then give *post-hoc* rationalisations

[3] The most recent account of Nietzsche's anti-realist arguments, specifically in support of an error theory, can be found in Robertson (2020): Ch. 3, 4.

[4] On the importance of Herder's influence on Nietzsche in this respect see Michael Forster (2017).

for. We can best explain substantive trends in moral judgements via a genealogy of how our sentiments have been socio-historically shaped for certain ends. For Nietzsche, our moral values, and the descriptive beliefs which underpin them, are deeply embedded into our psychology by a long history of selection pressures. *On the Genealogy of Morals* proffers a speculative history of how these pressures have favoured a 'herd instinct' in humans: a tendency to trample the exceptional in the interests of collective safety. If we can best explain moral judgements without recourse to objective moral properties, and we have reasons to think that our moral-belief-forming mechanisms are unreliable, then we have reason to doubt their existence.

Brian Leiter has interpreted these claims as constituting a distinctive argument from disagreement: "Nietzsche does, on this account, rely on *explanatory* considerations, but not with respect to our moral experiences *per se* but rather with regard to the phenomenon of moral disagreement (Leiter, 2014: 7). What is distinctive here, according to Leiter, is that Nietzsche focuses not on *folk* disagreement, but on the disagreement among alleged *experts* (namely: philosophers). Philosophers are in the business of finding the truth, and who often share background beliefs and practices. But, the argument goes, there is (embarrassingly) persistent and widespread disagreement even among these persons on foundational moral issues.

Leiter produces one passage from the *Nachlass* which appears to contain such an argument:

> It is a very remarkable moment: the Sophists verge upon the first *critique of morality*, the first *insight* into morality:—they juxtapose the multiplicity (the geographical relativity) of the moral value judgments [*Moralischen Werthurtheile*];—they let it be known that every morality can be dialectically justified; i.e., they divine that all attempts to give reasons for morality are necessarily *sophistical*—a proposition later proved on the grand scale by the ancient philosophers, from Plato onwards (down to Kant);—they postulate the first truth that a "morality-in-itself" [*eine Moral an sich*], a "good-in-itself" do not exist, that it is a swindle to talk of "truth" in this field. (*WP*, §428 [1888]).

However, textual credibility of the *Nachlass* aside, there are grounds for reservation about whether Nietzsche here makes the argument Leiter proposes. There are two concerns. First, the argument in the passage above looks very close to an argument which Nietzsche clearly *rejects* in the published works.

When considering the only "historians of morality" there have been so far, Nietzsche dismisses as "childish" how "they see the truth that among different nations moral valuations are *necessarily* different and then infer from this that *no* morality is at all binding" (*GS*, §345).[5] Earlier, when describing the historical and psychological acumen required for a reliable "study of moral matters", he writes how grand a project it would have to be to explain "the reasons for the differences between moral climates", and continues that "it would be *yet another job* to determine the erroneousness of all these reasons and the whole nature of moral judgements to date" (*GS*, §7 - emphasis mine).

While I am open to ways these passages may be reconciled, the published passages at least suggest more needs to be done to defend the argument from disagreement as genuinely representing Nietzsche's view. Leiter suggests that the argument from *WP*, §428 "has many analogues in the published corpus" (Leiter, 2014: 14). Nevertheless, the passages Leiter points to—*BGE*, §5, §186, §187, and *D*, Book I broadly—do not concern disagreement *per se*, but the broader point already discussed: that rational arguments to ground moral judgements are nothing but *post-hoc* justifications of affective attitudes. The second issue, then, is that the Nietzschean argument from disagreement Leiter proposes seems have little to do with *disagreement* itself. The claim that moral judgements are solely the product of the psychological constitution of those that assert them would run *even if there was total convergence in moral views*. Nevertheless, while there are grounds for doubt concerning whether Nietzsche deploys an argument from *disagreement*, the weight of evidence that he is a moral anti-realist is substantial.

1.2. Moral Judgements as Errors

So far, the supporting evidence substantiates only the *Ontological Thesis*. But this thesis is not sufficient for a moral error theory. A denial that moral judgements successfully capture objective moral facts is compatible with a rejection of its second essential component: the *Semantic Thesis* (i.e. that cognitivism is true). Perhaps moral judgements, such as 'stealing is wrong', do not express beliefs, but rather express *attitudes* such as approval or disapproval. If so, moral judgements do not express

[5] Attention to this passage in response to Leiter has also been given by Andrew Huddleston (2014): 329.

propositions, and consequently are incapable of being true or false. This view is non-cognitivism, and will be considered shortly. But first: are there good reasons to think Nietzsche endorses the *Semantic Thesis*?

Here a spectre of anachronism surfaces. It has been noted that Nietzsche did not have a sophisticated commitment to the semantics of moral discourse to straight-forwardly warrant labels such as 'cognitivist' or 'non-cognitivist', nor should he have been expected to, given that this meta-ethical dispute does not appear on the philosophical landscape in Europe until the mid-twentieth century (Leiter, 2000: 278–279, 2002: 137; Sinhababu, 2007: 264; Hussain, 2013: 412; Huddleston, 2014: 323; Silk, 2018). This is a reasonable claim to which I am sympathetic; the importance of which bears upon not just Nietzsche scholarship, but all appeals to pre-20th century philosophers in relation to contemporary ethical debates. I agree that Nietzsche, like any other 19th century European thinker, did not have a *sophisticated* account of the semantics of moral discourse. Nevertheless, such an account is not necessary. The idea that ordinary moral judgements do something *other* than report moral beliefs only became a sophisticated philosophical position after the introduction of non-cognitivist discourse. Hence, the spectre of anachronism does not threaten the error theory and non-cognitivism *equally*. As Sinhababu notes, error theory "has been around at least since Parmenides" (Sinhababu, 2007: 264, fn. 1), and it is only if we conceive of 'meta-ethics' in terms post-Ayer, Stevenson, and Hare, that 'sophistication' becomes exegetically problematic. The analogies Nietzsche draws between what he considers error-imbued non-moral judgements and moral discourse *at the very least* lend themselves to a consistent and unstrained reading of him as taking moral judgements to be systematically *false*.

Consider the following representative passage:

> *Astrology and what is related to it*. It is probable that the objects of the religious, moral [*moralisch*] and aesthetic experiences [*Empfindens*] belong only to the surface of things, while man likes to believe that here at least he is in touch with the heart of the world; the reason he deludes himself is that these things produce in him such profound happiness and unhappiness, and thus he exhibits here the same pride as in the case of astrology. For astrology believes the heavenly stars revolve around the fate of

> man; the moral man [*moralische Mensch*], however, supposes that what he has essentially at heart must also constitute the essence [*Wesen*] and heart of things. (*HH*, §4)

This passage is telling. Here Nietzsche is drawing an explicit parallel between astrology—a field in which the *propositions* intrinsic to its practice are *false*—and religion and morality. Astrology aspires to describe how the world is in a certain respect (namely, that the position of astronomical objects wholly determines human events and affairs), giving its practitioners corresponding *beliefs*. But since the world is not how astrologists describe it, these beliefs are false. Nietzsche takes the same process of error to apply *mutatis mutandis* to religious claims: "Moral judgment has this in common with religious judgement that it believes in realities which do not exist" (*TI*, 'Improvers': §1). That *morality* closely resembles these two mistaken fields in that its practitioner "deludes himself [*er täuscht sich*]" strongly suggests a *cognitive* component to Nietzsche's anti-realism.

The presence of both the *Semantic Thesis* and the *Ontological Thesis*—together of which are sufficient for a moral error theory—are also suggested in *Daybreak*:

> When man gave all things a sex he thought […] that he had gained profound insight:—it was only very late that he confessed to himself what an enormous error [*Irrthums*] this was […] — In the same way man has ascribed to all that exists a connection with morality and laid an *ethical significance* [*ethische Bedeutung*] on the world's back (*D*, §3)

He ends the passage that "One day" this view will "have as much value, and no more, as the belief in the masculinity and femininity of the sun has today" (*D*, §3). Again, Nietzsche appears to draw an analogy between a practice in which a property is ascribed [*beigelegt*] to *X,* resulting in a belief [*Glaube*] about *X*, yet since the property is absurd, the belief is false. Nietzsche repeats this strategy with seemingly cognitive language again in §103 when he writes that "it is *errors* which, as the basis of all moral judgment, impel men to their moral actions" (*D*, §103), this time drawing an analogy between morality and yet another discredited practice: "I deny morality as I deny alchemy, that is, I deny their premises: but I do *not* deny that there have been alchemists who believed in these premises" (*D*, §103).

A later passage from *The Gay Science* also strongly suggests an error theory. Addressing those apparently confident in the integrity of their moral principles, Nietzsche writes:

> …the *firmness* of your moral judgement could be evidence of your personal abjectness, of impersonality; your "moral strength" might have its source in your stubbornness—or in your inability to envisage new ideals. And, briefly, if you had thought more subtly, observed better, and learned more, you certainly would not go on calling this 'duty' of yours and this 'conscience' of yours duty and conscience. Your understanding *of the manner in which moral judgements have originated* would spoil these grand words for you, just as others grand words, like 'sin' and 'salvation of the soul' and 'redemption' have been spoiled for you (*GS*, §335)

The suggestion here appears to be that greater reflection upon the nature and origin of moral judgements would reveal an *error* inherent to them: that their status would be "spoiled" upon critical inspection *in the same sense* that previous concepts of importance—"sin", "salvation", and "redemption"—have been revealed to be misguided. If sin, for example, is understood in terms of the 'transgression of a divine law', but there is no divine law, then all judgments of the form 'X is sinful' are systematically false. The importance of 'sin' is thus "spoiled' insofar as we realise its fictionality, and consequently abandon our truth-apt beliefs about it. What indicates a moral error theory here is the suggestion that the individuals in question would *revise their beliefs* in the same way about *moral* concepts such as 'duty' *if* they had the ability to "envisage new ideals"; a requirement of which would be a greater depth of understanding into how moral judgements originate.

These passages are representative of Nietzsche's general tendency to conceive of moral judgements—like judgements about God, sin, astrology—as aiming to reporting objective *facts* about the world which lead to *beliefs* about the world. But since there are no 'moral facts', these beliefs are false: "We have measured the value of the world according to categories *that refer to a purely fictitious world*" (*WP*, §12 [1887-1888]). Before exploring this further, let us consider a competing interpretation flagged earlier.

2. Non-Cognitivism and Projectivism

Nietzsche has been interpreted as endorsing a form of non-cognitivism. The most sophisticated defence of this interpretation has been given by Maudemarie Clark and David Dudrick (2007). Clark and Dudrick's interpretation involves two relevant controversies for us. First, they hold that while Nietzsche held an error theory in *HH*, he came to endorse non-cognitivism in 1882 with the publication of *GS*. Second, they argue that the scope of his earlier error theory is *global*: at least in *HH*, Nietzsche is an error theorist about *all* evaluative judgements (Clark and Dudrick, 2007: 193). In the final section, I shall address the question of the scope of a Nietzschean error theory. But my present task is to determine whether Nietzsche does indeed reject or accept the *Semantic Thesis*. If Nietzsche is to be plausibly interpreted as holding an error theory about moral judgements, it must be the case that the interpretation of him as a non-cognitivist is false.

If *GS* suggests a change in Nietzsche's thought as significant as Clark and Dudrick suggest, the burden is on them (as they recognise) to provide adequate evidence. One of the major passages Clark and Dudrick cite in favour of a non-cognitivist reading is *GS*, §299; one of the many fragments which seek to collapse fundamental distinctions between ethical and aesthetic value. *GS*, §299 is entitled "*What one should learn from artists*", and suggests how our practice of valuing in the ethical domain can function in ways similar to aesthetic evaluation. Concerning the latter, Nietzsche asks "how can we make things beautiful, attractive, and desirable, for us when they are not? And I rather think that in themselves they never are". This firstly demonstrates a commitment to aesthetic anti-realism insofar as Nietzsche doubts that the aesthetic value of objects, persons, or states of affairs is ever wholly determined by their intrinsic properties. Given the context of the passage just described, Nietzsche would appear to be drawing the same conclusion of ethical value too, which sits consistently with other passages in which Nietzsche writes that "there is nothing good, nothing beautiful, nothing sublime, nothing evil in itself, but that there are states of soul in which we impose such words upon things external to and within us" (*D*, §210; cf. *HH*, §4), and that "Whatever has *value* in our world now does not have value in itself, according to its nature—nature is always value-less, but has been *given* value at some time, as a present—and it was *we* who gave and bestowed it" (*GS*, §301).

GS, §299 takes up the question of how we can "make" things have value. Nietzsche's answer, it appears, is that artists do this by manipulating the perceptions and perspectives people have of objects, which in turn alters our affective responses to them. In learning how to "partially conceal", and to see "through tinted glass", artists present natural phenomena in ways which elicit sentiment. Clark and Dudrick conclude that this passage conveys that, for Nietzsche, artists—and, Nietzsche hopes, ethicists too—can show us "how to evoke non-cognitive reactions, such as preferences and attitudes" (Clark and Dudrick, 2007: 203) as a means to genuinely *create values*; a project that makes little sense, they claim, if Nietzsche still maintained an error theory.

A second move Clark and Dudrick make to reinforce this non-cognitivist interpretation is to appeal to affinities between Nietzsche's claims and Hume's. In a familiar passage, Hume writes that as opposed to reason, taste "has a productive faculty, and gilding or staining all natural objects with the colours, borrowed from internal sentiment, raises, in a manner, a new creation" (Hume, 1998, App. 1.21). Hume's metaphor between colour and value to express the view that the sentiments *bestow* value upon objects is echoed by Nietzsche in multiple passages. sIn speaking of a potential science of morals, Nietzsche claims that "all that has given colour to existence still lacks a history", speaking in particular of the variety of "individual passions" which "have to be thought through and pursued through different ages, peoples…" (*GS*, §7). Section §139 of the same text—entitled "the colour of the passions"—contrasts respective moral outlooks by comparing the approach to the passions taken by St. Paul, who characterised them as "dirty, disfiguring", and the Greeks, who "loved, elevated, gilded [*vergoldet*], and deified them" (*GS*, §139).

On these grounds of affinity, Clark and Dudrick claim that "Nietzsche's meta-ethical position in *GS* is the basically Humean one that values are projections of passions and feelings" (Clark and Dudrick, 2007: 203). While I agree that drawing upon Hume to elucidate Nietzsche's position is fruitful, Clark and Dudrick presume Hume's meta-ethics reflects a straight forward non-cognitivism. But this is not the only interpretation of his view on offer: 'projectivism' about value is naturally compatible with a moral error theory, and there are good reasons to suppose Nietzsche takes a similar line.

Projectivism, in its broadest formulation, is a commitment to two theses:

(1) We experience moral properties as mind-independent features of the world.

(2) This experience has its origin in affective attitudes (e.g. sentiments of approval or disapproval) which are causally responsible for it.[6]

So far, this understanding of projectivism is neutral between moral realism and moral anti-realism, since (2) simply offers a causal mechanism which explains how it is our psychology functions to give us the phenomenal experience described in (1). Anti-realist versions of projectivism are entailed only with an additional thesis:

(3) In fact, moral properties do not exist mind-independently.

This combination of theses is compatible with at least three distinct and competing anti-realist views: (a) non-cognitivism; (b) moral error theory; (c) subjectivism (i.e. the view that moral judgements express propositions but the truth-conditions for such propositions are mind-dependent).

The non-cognitivist interpretation of Nietzsche can thus be understood as a form of projectivism: we 'colour' objects with value our affects cast upon them, and our *experience* of them is as if they are mind-independent.[7] However, projectivism is a thesis that is congenial to the error theory. To get to a moral error theory, (1)-(3) must be endorsed alongside a further step:

(4) When we utter sentences such as 'X is morally right' or 'X is morally virtuous' etc, we purport to ascribe properties to X. But our belief in such properties is systematically mistaken, making such utterances false.

So the error theorist can accommodate the understandable concerns of Clark and Dudrick's non-cognitivism. This becomes pertinent with Clark and Dudrick's appeal to affinities between Nietzsche and Hume to ground a non-cognitivist interpretation. They assume that Hume is a non-cognitivist because of his apparent endorsement of (2). But as we have seen, (2) can be constitutive of a projectivist moral error theory. Identifying Hume's meta-ethical position, like Nietzsche's, faces a

[6] Richard Joyce labels this "minimal projectivism". See Joyce (2009). In what follows I broadly follow his approach.

[7] It might seem that non-cognitivism must be committed to a denial of (1) in order to distinguish itself from an error theory, but this is not the case. That we experience *X* has having a certain character does not *entail* that we believe *X* to have that character.

number of interpretive difficulties. Nevertheless, it is no surprise that Hume has been interpreted along projectivist lines (Mackie, 1980; Stroud, 1993; Kail, 2007; Hussain, 2012; Olson, 2014).

Let us reconsider the famous passage from Hume quoted above in which he describes how taste (i.e. the passions) 'gilds and stains' objects with 'colour' derived from sentiment. Clark and Dudrick rely upon this passage as evidence of Hume's non-cognitivism (supposedly echoed by Nietzsche). But we can now see how this can naturally be built into an error theory about moral judgements. It is true that the passage does not explicitly mention *error*, but this is not surprising. Read in its context, it concerns not moral semantics or the existence of moral properties, but the psychological workings of taste as opposed to reason with respect to *motivation*. Other equally famous passages from Hume have been offered which *do* suggest error in projection. For instance:

> ...the mind has a great propensity to spread itself on external objects, and to conjoin with them any internal impressions, which they occasion, and which always make their appearance at the same time that these objects discover themselves to the senses. Thus as certain sounds and smells are always found to attend certain visible objects, we naturally imagine a conjunction, even in place, betwixt the objects and qualities, tho' the qualities be of such a nature as to admit of no such conjunction, and really exist no where (Hume, 1985: 217)[8]

Once again, Nietzsche's claims of projection are along the same lines. In *HH* he suggests that "for thousands of years" our moral, religious, and aesthetic judgements have been born from "blind inclination, passion, or fear", thus we have "indulged ourselves fully in the bad habits of illogical thought". But while, as a result, "this world has gradually *become* so strangely colourful", this is because "we have been the painters: the human intellect allowed appearance to appear, and projected [*hineingetragen*] its mistaken conceptions onto things" (*HH*, §16). He writes later that there is no good in-itself, "but that there are states of soul [*Seelenzustände*] in which we *impose* [*belegen*] such words upon things external to and within us" (*D*, §210 - emphasis mine), and in a notebook passage from his mature period suggests why:

[8] For additional criticisms of Clark and Dudrick's interpretation of Hume see Hussain, (2012): 128-131.

> All the values by means of which we have tried so far to render the world estimable for ourselves […] all these values are, psychologically considered, the results of certain perspectives of utility, designed to maintain and increase human constructs of domination—and they have been falsely *projected* [*projicirt*] into the essence of things (*WP*, §12 [1887-1888])

These passages, and the others I have considered here, suggest Nietzsche's endorsement of projectivist theses (1)-(3), and *at the very least* are *consistent* with (4), which Clark & Dudrick must deny. For these reasons, the claim that Nietzsche abandoned the cognitive component of his error theory cannot be substantiated by an appeal to a projectivist view. However, this still leaves open how value creation is possible. I now turn to this issue in addressing the *scope* of Nietzsche's error theory.

3. The Scope of Error

So far, I have interpreted Nietzsche to hold a restricted error theory, that is: an error theory about *moral* judgements in particular. However, the projectivism just considered appears to equally apply more broadly to non-moral value judgements (e.g. aesthetic judgements), and in many passages Nietzsche seems to endorse a wider scope of error in this domain. In *HH*, he considers "the *necessary* injustice in every For and Against" (*HH*, Preface: §6; cf. *D*, §210; *GS*, §301). In many of the passages discussed earlier, Nietzsche explicitly includes *aesthetic* judgements as suffering from the same errors (*D*, §210; *GS*, §299, §301). Hence, it is no surprise that Nietzsche has been read as error theorist about all evaluative judgements, as Nadeem Hussain's influential interpretation does. Hussain claims that for Nietzsche "all claims of the form 'X is valuable' are false"' (Hussain, 2007: 159). Charles Pigden defends a similar error theoretic reading of Nietzsche (Pigden, 2007: 443-446), as do Clark and Dudrick in their exegesis of *HH* (Clark & Dudrick, 2007: 200-201), as does Reid Blackman's recent account (Blackman, 2020).

The problem is that Nietzsche very often appears to emphatically make value judgements or, at the very least, make claims which *presuppose* the existence of genuine values. From the standpoint of a broadly perfectionist axiology which centres on achievement, creativity, and freedom, Nietzsche frequently demands others to *reconsider the real worth* of contemporary values such as compassion,

happiness, altruism, equality, and so forth. In *GM*, Nietzsche sets out to investigate the "value of these values" (*GM*, Preface: §6) out of a *need*: they might in fact be *harmful* or otherwise inimical to manifesting *genuine* value. If Nietzsche maintains an error theory about moral judgements as suggested, how can he make such claims without embodying a blatant inconsistency?

In order to answer this question and in turn clarify the scope of his error theory, two crucial distinctions must be made. The first distinction is in Nietzsche's use of the term 'morality'. Across his corpus, Nietzsche's preferred terms—*Moral*, *Moralität*, and *Sittlichkeit*—are applied often interchangeably to refer to two phenomena: (1) any system of values, beliefs, practices endorsed by a society or individual; (2) a *particular* system of values, beliefs, and practices, namely those inherent to the Judeo-Christian worldview. Briefly, instead of a monolithic list of substantive prohibitions and commands, this narrower sense of morality is perhaps best thought of as a broad family of normative commitments (including the descriptive views they depend upon: e.g. about agency and free will) which typically takes pity/compassion, equality, happiness and altruism to be of intrinsic value, and takes the status of this value to be both objective and unconditional. Following Leiter (2002), I shall refer to (2) as 'morality in the pejorative sense' (MPS).

This distinction is most clearly expressed in *BGE*, §202:

> *Morality is in Europe today herd-animal morality*—that is to say, as we understand the thing, only *one* kind of human morality beside which, before which, after which many other, above all *higher*, moralities are possible or ought to be possible (*BGE*, §202)

Nietzsche explicitly draws this conceptual line in the sand soon after when writing that "Beyond Good and Evil. — At least this does not mean 'Beyond Good and Bad'" (*GM*, I: §17). These claims suggest that Nietzsche intends an error theory about MPS, but not about value broadly. Nevertheless, his projectivism appears to offer no non-arbitrary reason for excluding certain value judgements from the boundaries of error.

I propose a second distinction pertinent to this interpretative puzzle; a distinction which will again draw upon Hume. Jonas Olson interprets Hume as having a two-fold meta-ethic: first is an "account of

actual or vulgar moral thought and talk, that is to say, the moral thought and talk of ordinary people"; second is an account of "how actual or vulgar moral thought and talk could be reformed so as to no longer involve error" (Olson, 2014: 21). In Olson's view, the former is characterised by a projectivist moral error theory, and the latter subjectivism. Olson provides credible evidence for this distinction. To take but one example, he quotes Hume's assertion in *A Treatise on Human Nature* that "Vice and Virtue […] may be compar'd to sounds, colours, heat, and cold, which, according to modern philosophy, are not qualities in objects, but perceptions in the mind", and that "this discovery in morals, like that other in physics, is to be regarded as a considerable advancement of the speculative sciences" (Hume, 1985: 520-521). The relevant point here that Olson exploits is that Hume's characterisation of this analogy as a "discovery" and a "considerable advancement" in human knowledge strongly suggests that projectivism is *not* what ordinary people think is going on in moral discourse. This distinction allows for Hume to hold a moral error theory in one sense, and an alternative—in Olson's interpretation: a subjectivism—for a revisionary meta-ethic.

Whether in fact Hume holds such a meta-ethical view is an interesting question in-itself, but strictly speaking irrelevant for my purposes here. I claim only that this type of dialectical manoeuvre finds its analogue in Nietzsche. Crucially, this distinction between ordinary evaluative views and the views of a revolutionary elite is not an *ad hoc* invention to solve the current problem, but is ubiquitous in Nietzsche's texts after 1878. A persistent theme is the need for certain persons to 'create values', or, as Nietzsche puts it: to "*fashion* something that had not been there before: the whole eternally growing world of valuations, colours, accents, perspectives, scales, affirmations, and negations" (*GS*, §301). Although this passage naturally has an anti-realist quality to it, at present it is unspecific about what value creation actually involves. Before returning to some suggestions, I wish to draw attention to Nietzsche's repeated assertions that it is a practice *markedly different* from how the majority understand and experience value.

When Nietzsche envisions a "purification of our opinions and valuations and to the *creation of our own new tables of what is good*" (*GS*, §335) and speaks of "we who think and feel at the same time" (*GS*, §301) in creating values, the collective "we" and "our" being referred to is *not* the human *per se*, but a

specific (and minority) group who have the strength for reconsidering the nature of value. Nietzsche is clear on this point. In *HH* he distinguishes free spirits as those able to forge and control their affective attitudes in determining values: "You had to gain power over your For and Against, and learn how to hang them out or take them in, according to your higher purpose" (*HH*, Preface: §6). More explicitly in *BGE*, Nietzsche refers to those spirits "strong and original enough to make a start on antithetical evaluations and to revalue and reverse 'eternal values'" (*BGE*, §203). Strength is required because creating new values is *difficult*: "To seize the right to new values—that is the most terrible proceeding for a weight-bearing and reverential spirit (*Z,* I: 'Of the Three Metamorphoses'). The creation of new values will be difficult for a number of reasons. For example, creating new values, in opposing traditional values, will typically induce conflict with other members of society. A psychological strength will thus be required to bear (1) the *solitude* and *isolation* (e.g. *GS*, §296, §297; *BGE*, §212); (2) the responsibility for causing *harm* to adherents to old values (e.g. *GS*, §311, §325), which will likely be a result of challenging established norms.

Contrast this with the weak, which Nietzsche describes as having the "inability to envisage new ideals", and explicitly claims that this constitutes at least part of the explanation of the confidence in, or "*firmness* [*Festigkeit*]" (*GS*, §335) of, their moral judgements. The free spirit or "genuine philosopher" (*BGE*, §211), by contrast, *does* have this ability in virtue of their strength:

> The strongest and most evil spirits have so far done the most to advance humanity […] they reawakened again and again the sense of comparison, of contradiction, of the pleasure of what is new, daring, untried […] usually by force of arms, by toppling boundary markers, by violating pieties—but also by means of new religions and moralities (*GS*, §4)

Let us call the moral discourse of ordinary people—or to use Nietzschean terminology: the herd—*conventional evaluative practice* (CEP). Let us call the moral discourse of the free spirits who are able to 'create values' *revisionary evaluative practice* (REP). On my reading, Nietzsche is committed to holding that CEP is characterised by a projectivist error theory, whereas REP is not. Under the category of CEP would be MPS: Christian morality and its secular derivatives claim to get to 'the heart of things'; to make objective reports about how the world is in-itself. Moreover, it claims to do this with a unique and

universal authority. After distinguishing between MPS and the possibility of genuine normativity, Nietzsche writes that "against such a 'possibility', such an 'ought', this morality [MPS] defends itself with all its might: it says, obstinately and stubbornly, 'I am morality itself, and nothing is morality besides me!'" (*BGE*, §202). The important point here is that other systems of morality in the broad sense may also be characterised by an error theory if they commit the same meta-ethical mistakes of MPS. This holds even if those systems of morality qualify as 'higher' on a Nietzschean analysis, for Nietzsche clearly denies that *false* beliefs are always devoid of value and ought to be discarded (e.g. *BGE*, §4). Nevertheless, some possible sense of normativity does not *necessarily* have to be infected with error. This opens up the space for REP, which Nietzsche calls for "genuine philosophers" to explore. But what *is* REP by Nietzsche's account?

Hussain's error theory, because it includes *all* evaluative discourse, is combined with a form of *moral fictionalism* in order to solve the puzzle of how Nietzsche continues to make value judgements. Moral fictionalism holds that when *A* values *X*, *A* regards *X* as valuable in itself while knowing that in fact *X* is not valuable in itself. Hussain describes fictionalists as valuers engaged in "a simulacrum of valuing" (Hussain, 2007: 158). In order to solve the interpretive puzzle, Hussain's understanding of REP would be the following:

> Nietzsche's recommended practice is a form of make-believe or pretence. Nietzsche's free spirits pretend to value something by regarding it as valuable in itself while knowing that in fact it is not valuable in itself (Hussain, 2007: 170)

Hussain's fictionalism has been criticised for a various reasons. For example, it is difficult to reconcile with Nietzsche's repeated emphasis on the value of *honesty* embodied by 'free spirits' (Silk, 2015: 273-274);[9] it does not account for Nietzsche's claims that value creation entails *making* things valuable which previously were not, and not simply pretending *as if* they were valuable (Clark & Dudrick, 2007; Silk, 2015: 273); and that even if it could account for value *creation*, it is not a plausible account of *revaluation*, which Nietzsche calls for (Thomas, 2012). I am sympathetic to these criticisms.

[9] Consider, for example: "How much truth can a spirit *bear*, how much truth can a spirit *dare*? That became for me more and more the real measure of value […] error is cowardice…Every acquisition, every step forward in knowledge is the *result* of courage (*EH*, Foreword: §3; cf. *BGE*, §39).

Although a comprehensive critique of the fictionalist interpretation is not possible here, I wish to consider one of Hussain's justifications for his fictionalist interpretation.

Hussain's strategy is to show that alternative readings of Nietzsche's views fail to do justice to his claims, and so fictionalism is the only plausible candidate left standing. One alternative he criticises is subjective realism, the view that moral judgements express beliefs, but the moral properties to which they refer are mind-dependent. On this view, values are grounded in subjective attitudes of valuing, but nonetheless *also* gain some authoritative standing in the world for the creator—and possibly the community to which they belong—once 'created'. Hussain writes that the subjectivist interpretation cannot account for the passages in which Nietzsche appears to endorse an error theory about all evaluative judgements (moral and non-moral):

> …it seems that a subjective realism about non-moral evaluations would have trouble with such passages. After all, if indeed evaluative claims have the proposed subjective truth-conditions, then they do not get the world wrong. They do not seem to involve any essential intellectual loss (Hussain, 2007: 162)

However, Hussain's concern here dissolves once we deploy the distinction between CEP and REP that I have defended. With this conceptual apparatus, we can hold that Nietzsche *does* think ordinary evaluative discourse in moral *and* aesthetic domains essentially embodies error, but that a revisionary form of valuing reserved for an elite few need not do so. The problem then, is that Hussain's interpretation has Nietzsche understanding evaluative discourse as *necessarily* involving error. But as we have seen, there are good reasons to deny this claim: Nietzsche thinks evaluative discourse—in which things which have no value in themselves are genuinely *made* valuable as opposed to pretending so—is salvageable, if only for a minority of strong individuals. This opens up a gap for competing interpretations, such as a subjectivism, to exploit.

But the subjectivist interpretation has also come under fire from Clark & Dudrick, who find it implausible that *X* might be good *just because* I value *X*. They write:

> There is no doubt that Nietzsche recognises certain virtues: e.g., loyalty, honesty, courage. But it is no more plausible that courage is good or admirable because people admire it than that murder is wrong because people disapprove of it (Clark & Dudrick, 2007: 205)

I agree with Clark & Dudrick that a meta-ethical view such as *that* would be implausible. However, these are poor grounds to reject a subjectivist interpretation of REP, since the version they reject is particularly crude. There are sophisticated versions of subjectivism which cannot be dismissed so easily. It will be worth briefly mapping a potential version which, while requiring further defence than I can provide here, helpfully elucidates the available space for genuine value creation, given a local error theory.

Nietzsche has been read as endorsing various sophisticated versions of subjectivism (e.g. Anderson, 2005, Silk, 2015). One such version—which REP could be understood in terms of—is constructivism. Broadly expressed, the form of constructivism relevant here is the view that evaluative facts are grounded solely in facts about an agent's evaluative *attitudes*. What would make a normative judgment *correct* is that it coheres with the relevant agent's evaluative attitudes. More precisely, constructivism holds that genuine evaluative standards which establish the relevant set of evaluative facts are constituted by their emergence from a distinctive practice. Moreover, that it is these facts that make our evaluations true or false.

At present, this of course leaves open which distinctive practice grounds a particular set of evaluative facts. There are many forms of constructivism which provide competing answers. Kantian versions of constructivism, for example, understand the nature of normative truths in terms of considerations about the basic features of *rational agency*. On this view, reasons for being moral are derived from our nature as rational agents. Insofar as moral obligations are justified in terms of these rational requirements, they are universally and categorically binding for all who fall within the class of 'rational being'.

Alex Silk (2015) has developed a distinctively Nietzschean constructivism, according to which evaluative facts are grounded in facts about the evaluative attitudes of 'genuine philosophers' or 'free spirits' who are capable of taking on a variety of *perspectives* of phenomena which in turn refines their

judgements about them. Silk's strategy is to harness Nietzsche's 'perspectivism'—i.e. the view that all knowledge is 'interested' and perspectival, and that greater truth can be attained to the extent that one takes on multiple perspectives via a process of critical examination of our interests and affects—in service of his meta-ethical project. Concerning genuine "objectivity", Nietzsche writes that it ought to be…

> …understood not as "contemplation without interest" (which is a nonsensical absurdity), but as the ability *to control* one's Pro and Con and to dispose of them, so that one knows how to employ a *variety* of perspectives and affective interpretations in the service of knowledge (*GM*, III: §12, cf. *HH*, Preface: §6)

Applied to the evaluative domain, values are, on this view, "properties of one's own perspective, but not 'merely' one's own perspective in any sense to be disparaged" (Silk, 2015: 258). This version of constructivism does not maintain that agents can simply *choose* to value something and that would make it valuable. Rather, genuine valuing requires *fashioning* and *forging* one's affects by way of rigorous self-examination and careful analysis of natural and social selection pressures (hence, the essential requirement of genealogical investigation). This is what it means to *create* values.

I have already given two reasons why value creation is *difficult*, but the taking on of multiple perspectives in this way provides a third reason, characterising value creation as a genuine *achievement*, and (like artistic creation) probably not open to all. Because value creation is possible only for 'higher types' with the strength to do so, this makes Nietzschean constructivism distinct from Kantian versions insofar as the former lacks the egalitarianism built into the universality of the latter.

As well as accounting for how value creation is possible and why Nietzsche thinks it is available only to a minority, another significant advantage to the constructivist interpretation is that accounts for how values are legitimately *demanding*, and allow for agents to sometimes be mistaken about which things are good. Constructivism offers an account of how evaluative practice can exhibit the genuine *normative force* which Nietzsche repeatedly emphasises values to have, while at the same time denying that anything has value 'in-itself' (i.e. independently of attitudes), thus keeping the *Ontological Thesis* in tact. A subjectivist interpretation of REP then, does not have to take a crude and implausible form: values can be created, and hence be dependent on evaluative attitudes, but maintain genuine normative force: "no

longer the humble expression, 'everything is *merely* subjective' but 'it is also *our* work!—let us be proud of it!" (*WP*, §1059 [1884]; cf. *GS*, §335).

While a comprehensive analysis of the constructivist interpretation is beyond the scope of this paper, by raising it as a possibility I hope to open up greater space for understanding how REP is to be understood; space which I have argued is too hastily closed off by both Hussain and Clark & Dudrick. The two distinctions proposed then, allow Nietzsche to maintain a *local* moral error theory—where 'moral' includes MPS—while simultaneously making genuine value judgements in a revised sense. This gives us a vantage point for determining how Nietzsche's meta-ethical position interestingly differs from contemporary formulations. If the interpretation given so far is sound, Nietzsche can be read as holding that there *are* evaluative properties, but that the nature of these properties is drastically *different* from what the majority of humans—'the herd'—experience them as. The majority experience evaluative properties as mind-independent features of the world, but since properties *of that kind* do not exist, their beliefs about them are systematically false.

Conclusion

This paper has reconsidered the grounds for attributing to Nietzsche an error theory about moral judgements. I have argued that there are plausible reasons for interpreting Nietzsche as a projectivist moral error theorist, yet in a significantly restricted sense relative to Nadeem Hussain's influential reading. This difference emerges from exploiting two distinctions: (1) 'MPS' and 'morality in a broad sense'; (2) *conventional evaluative practice* and *revisionary evaluative practice*. Drawing these distinctions allows for a consistent error theory to be maintained while also making genuine value judgements. These value judgements, I have suggested, are most parsimoniously explained via a subjective realist framework while remaining faithful to Nietzsche's texts. I have suggested that, like Hume, Nietzsche's error theory is projectivist in nature. Hence, I think Clark and Dudrick are right to draw affinities with Hume in helping elucidate Nietzsche's position. However, as I have argued, they draw the analogy too hastily. At the very least, the availability of a more sophisticated form of subjective realism than the one used to reject an error-theoretic interpretation shifts the burden of proof back onto proponents of the non-cognitivist reading.

Abbreviations

Works by Nietzsche are cited by section using the following translations (though modified where I have felt it appropriate to do so):

BGE = *Beyond Good and Evil*, trans. R.J. Hollingdale, 1990

D = *Daybreak*, trans. R.J. Hollingdale, 1997

EH = *Ecce Homo*, trans. R.J. Hollingdale, 1986

GM = *On the Genealogy of Morals*, trans. W. Kaufmann and R. J. Hollingdale, 1989

GS = *The Gay Science*, trans. Walter Kaufmann, 1974

HH = *Human, All Too Human*, trans. R.J. Hollingdale, 1996

TI = *Twilight of the Idols*, trans. R.J. Hollingdale, 1968

Z = *Thus Spoke Zarathustra*, trans. Walter Kaufmann, 1954

WP = *The Will to Power*, trans. Walter Kaufmann and R.J. Hollingdale, 1967

Bibliography

Anderson, Lanier, (2005), "Nietzsche on Truth, Illusion, and Redemption", *European Journal of Philosophy*, 13(2): 185–225.

Blackman, Reid, (2020), "Nietzsche's Metaethics: Fictionalism for the Few, Error Theory for the Many", in Paul Katsafanas (ed.), *The Nietzschean* Mind, Routledge.

Clark, Maudemarie and Dudrick, David, (2007), "Nietzsche and Moral Objectivity", in Brian Leiter and Neil Sinhababu (eds.), *Nietzsche and Morality*, OUP: 192-226.

Forster, Michael, (2017), "Nietzsche on Morality as a 'sign language of the affects'", *Inquiry*, 60:1-2: 165-188.

Huddleston, Andrew, (2014), "Nietzsche's Meta-axiology: Against the Sceptical Readings", *British Journal for the History of Philosophy*, 22:2: 322-342.

Hume, David, (1985), *A Treatise of Human Nature,* Ernest Mossner (ed.), Penguin Classics.

—(1998), *An Enquiry Concerning The Principles of Morals*, T.L. Beauchamp (ed), Clarendon Press.

Hussain, Nadeem, (2007), "Honest Illusion: Valuing for Nietzsche's Free Spirits", in Leiter and Sinhababu: 157-191.

— (2012), "Nietzsche and Non-Cognitivism", in Simon Robertson and Christopher Janaway (eds., *Nietzsche, Naturalism & Normativity*, OUP: 111-132.

— (2013), "Nietzsche's Meta-ethical Stance", in the *Oxford Handbook of Nietzsche*, Ken Gemes and John Richardson (eds.), OUP: 389-414.

Joyce, Richard, (2001), *The Myth of Morality*, CUP.

— (2006), *The Evolution of Morality*, MIT Press.

— (2009), "Is Moral Projectivism Empirically Tractable?", *Ethical Theory and Moral Practice*, Vol. 12: 53-75.

Kail, Peter, (2007), *Projection and Realism in Hume's Philosophy*, OUP.

Leiter, Brian, (2014), "Moral Skepticism and Moral Disagreement in Nietzsche", Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics*, Vol. 9, OUP.

— (2002), *Nietzsche on Morality*, Routledge.

— (2000), "Nietzsche's Metaethics: Against the Privilege Readings", *European Journal of Philosophy*, 8:3: 277-297.

Mackie, John, (1977), *Ethics: Inventing Right and Wrong,* Penguin Books.

— (1980), *Hume's Moral Theory*, Routledge.

Olson, Jonas, (2014), *Moral Error Theory: History, Critique, Defence*, OUP.

Pigden, Charles, (2007), "Nihilism, Nietzsche and the Doppelganger Problem", *Ethical Theory and Moral Practice*, Vol. 10, No. 5, Moral Skepticism: 30 Years of Inventing Right and Wrong: 441–456.

Silk, Alex, (2015), "Nietzschean Constructivism: Ethics and Metaethics for All and None", *Inquiry*, 58:3: 244-280.

— (2018), "Nietzsche and Contemporary Meta-Ethics" in Paul Katsafanas (ed.) *Routledge Philosophical Minds: The Nietzschean Mind*, Routledge.

Robertson, Simon, (2020), *Nietzsche and Contemporary Ethics*, Oxford: Oxford University Press.

Sinhababu, Neil, (2007), "Vengeful Thinking and Moral Epistemology", in Leiter and Sinhababu (eds.).

— (2015), "Zarathustra's Metaethics", Canadian Journal of Philosophy, Vol. 45, No. 3: 278–299

Streumer, Bart, (2017), *Unbelievable Errors: An Error Theory About All Normative Judgements*, OUP.

Stroud, Barry, (1993), "'Gilding' or 'Staining' the World With 'Sentiments' and 'Phantasms'", *Hume Studies* 19: 253-72.

Thomas, Alan, (2012), "Nietzsche and Moral Fictionalism", in Simon Robertson and Christopher Janaway (eds.), *Nietzsche, Naturalism & Normativity*, OUP.

Williams, Bernard, (1980), "Internal and External Reasons" reprinted in his *Moral Luck*, CUP.