

ARISTOTELIAN NATURALISM VS. MUTANTS, ALIENS, AND THE GREAT RED DRAGON

Scott Woodcock

ABSTRACT

In this paper I present a new objection to the Aristotelian Naturalism (AN) defended by Philippa Foot. I describe this objection as a *membership* objection because it reveals the fact that AN invites counterexamples when pressed to identify the individuals bound by its normative claims. I present three examples of agents for whom the norms generated by AN are not obviously authoritative: mutants, aliens, and the Great Red Dragon. Those who continue to advocate for Foot's view can give compelling replies to the first two of these examples, but their replies drive the view into an unwelcome result when it faces the last example. I conclude that the concept of being human, on which AN crucially depends, is not as straightforward as Foot's advocates presume.

I. INTRODUCTION

Naturalistic forms of Aristotelianism remain influential in contemporary moral philosophy. In particular, the type of Aristotelian Naturalism defended by Philippa Foot in *Natural Goodness* (2001) continues to inspire sympathetic interpretations from her supporters (e.g. Brown 2004; Hacker-Wright 2009a, 2009b, 2012, 2013a, 2013b; Hendley 2009, 2015; Teichmann 2011; Lott 2012a, 2012b, 2015).¹ Foot's view drew extensive criticism from those skeptical of teleological accounts of natural phenomena serving as standards for human normativity (Fitzpatrick 2000; MacIntyre 2002; Slote 2003; Copp and Sobel 2004; Andreou 2006; Millum 2006; Woodcock 2006; Millgram 2009; Lewens 2010). Nevertheless, Foot's advocates clarify her arguments in critical respects, and they offer insightful replies to counterexamples

that otherwise appear devastating to the view. The result is a spirited debate between those continuing to support Foot's naturalism and those still suspicious of its normative appeal.

In this paper I present a new objection to Foot's view. I call it a *membership objection* because it reveals that Aristotelian Naturalism (AN) encounters problems when it attempts to identify the individuals bound by its normative claims. I present three counterexamples of individuals for whom the norms of AN are not obviously authoritative. The supporters of AN can give compelling replies to the first two of these examples, but their replies drive the view into an unwelcome trilemma when it faces the third example. In this last type of example, AN fails to provide norms of practical rationality that are *categorically* binding for ordinary human agents. Thus, my aim will be to demonstrate that the concept of being *human*—a concept on which AN crucially

depends—is not as straightforward as its advocates presume. I argue that this view is subsequently left without a principled basis to establish authoritative norms for the intended membership of its moral community.

2. FOOT'S ARISTOTELIAN NATURALISM

The principal thesis of AN is that normative claims of goodness and defect are determined by natural facts that describe how individuals ought to live as representative members of their species. Just as we make non-instrumental claims about the goodness of living things like oak trees according to whether they exemplify the characteristic features of their particular life cycle (e.g. solid roots, many acorns, etc.), goodness for humans is defined according to natural history descriptions of how humans tend to flourish as representative members of our distinct life cycle (Foot 2001, 25–37). It remains a point of contention whether Foot intends to merely draw an analogy between human goodness and the goodness of other living things, or whether she is proposing that judgments of human goodness are literally a subset of a more general set of judgments regarding natural goodness (Hursthouse 2012, 177; Hendley 2015). Nevertheless, Foot emphasizes that the logical form of our evaluations of human goodness and defect is no different than the form of these evaluations for other living things.

Foot's work draws heavily on the work of Michael Thompson (1995), who continues to provide helpful clarifications of how attributions of goodness are best interpreted from the AN perspective (2003, 2004, 2008). Thompson underscores two key aspects of the natural history judgments that ground AN. First, he points out that the set of statements that specify the proper functioning for a particular species—what he refers to as “Aristotelian categoricals”—need not be true for individual members of a species. In fact, they need not even be true for a majority of the

species, as odd as that may seem. A helpful example Thompson provides is an umbrella jellyfish that can be described as having a certain number of tentacles and developing from an egg to a mature adult specimen (2004, pp. 48–51). It may be that the average number of tentacles for existing members of the species has never been exactly equal to this number, and hundreds more eggs will fail to develop than will survive to live out a full life cycle for the species. Yet we can nevertheless speak meaningfully about the set of norms that describe archetypal members of this species. Thus, characteristic claims like, “An umbrella jellyfish has 140 tentacles” are not refuted by empirical observations of jellyfish with more or fewer tentacles, nor by statistical population averages. Instead, the claim is a natural history description of what it is to flourish according to the natural norms for representative members of this species.

Second, Thompson emphasizes that the natural history judgments proposed by AN are not empirical claims like those provided by evolutionary biology. Instead, these judgments are generated by the unique logical form that is *presupposed* when one performs the interpretive task of organizing raw, empirical observations into the coherent classification of a particular life form (Thompson 1995, pp. 284–88). Thus, from the AN perspective, the claim that a wolf ought to hunt in coordination with other pack members is not straightforwardly determined by empirical observations of wolf behavior; rather, it expresses a normative standpoint we adopt when we assemble a profusion of otherwise chaotic sensory information and *represent* a wolf as a living organism with the vital functions it needs to survive and reproduce in its own unique way. Natural history judgments are therefore distinct from the data supplied by biological sciences, because the judgments involve representations of functional organization that we put forward when we delineate patterns of natural normativity for

living creatures (Hacker-Wright 2009, pp. 316–17; Lott 2012a, pp. 353–58).

Finally, as one ought to expect from a version of Aristotelianism, Foot claims that when it comes to humans the norms that specify goodness and defect for our life cycle are grounded in our distinctive capacity for reason (2001, pp. 66). Evaluations of how a human ought to live as a member of our species are confined to dispositions of one's rational will, for this above all else exemplifies a unique pattern of normativity for our life form in the same way colorful plumage is a distinctive pattern of normativity for male peacocks. Both evaluations of goodness share a similar logical form, but norms of natural goodness and defect for humans are *sui generis* in the way they refer to dispositions of a rational will. Thus, the defining feature of our species sets us apart from all other species of animals and plants, yet this feature is part of a general system of normativity that applies to all living organisms in the natural world.

3. MUTANTS

My first prospective counterexample for AN begins with the observation that species are not fixed populations of identical members. Instead, considerable degrees of variation exist within the boundaries of individual species, and establishing a coherent view of species taxonomy has become a notoriously difficult task in the philosophy of biology (e.g. Hull 1965a, 1965b; Sober 1980; Kitcher 1984; Ereshefsky 2001, 2009, 2010; Okasha 2002; Wilkins 2009). We may find it convenient to rely on common-sense notions of species membership, but these notions retain enduring traces of essentialism that cannot be reconciled with how difficult it is for biology to identify common traits that unify the members of individuated species. When the complexity of natural phenomena is acknowledged, there are, surprisingly, no clear traits that are present in all and only

the members of our common-sense species classifications.²

As critics of AN like Alasdair MacIntyre have noted, this variation in mixed populations allows for the possibility that some types of immoral behavior (e.g. defection from norms of promise-keeping) are consistent with enduring subsets of the human species (2002, pp. 626–27).³ The reply to this objection from AN advocates is to emphasize the difference between empirical observations derived from biological sciences and the non-observational judgments that specify Aristotelian life forms. By highlighting this difference, AN avoids being drawn into the complex empirical details of variation within the members of particular species.⁴

(i) *The Mutant Puzzle*

The counterexample to AN that I present, however, appeals to a different type of intraspecies variation. Rather than emphasizing variation in mixed populations at a particular time, consider the variation of species as they change over time. Existing variation in a species allows some of the mutants in an initial population to emerge from their selective environment as the new representative members of that species. As Elijah Millgram notes, the variation that allows for new traits, and in some cases speciation, is itself a trait that must be included in Aristotelian categoricals that specify the logical forms of species (563). Yet this creates a counterexample for AN: what to say about mutants who exhibit a trait at time t_1 that will later reach fixation in the population at time t_2 and become a distinctive feature of the newly evolved species? It is counterintuitive to maintain that the mutant is *defective* at time t_1 when an identical member of the species will later exhibit natural goodness at time t_2 in virtue of the very same trait that eventually reaches fixation.

For example, the disposition to issue alarm calls is a vital trait in populations of meerkats

who rely on sentinel warnings to identify predators (Hollén and Radford 2009), and it would be odd to describe the new mutants at each stage in the phylogeny of this disposition as defective members of their species: this despite the fact that each of these mutants would be exhibiting atypical (and not yet obviously adaptive) behavior at each successive stage of development of the species. The looming problem for AN is that this applies to *every* trait currently exhibited by species. For every peacock with colorful plumage there existed a prior mutant with this same plumage who was not representative of his species at that earlier stage of phylogeny. For every human with a disposition for sympathy there was some prior mutant whose increased capacity for sympathy was not consistent with the traits that characterized the human life form at that prior evolutionary stage.⁵ Hence, the challenge for AN is to offer some principled way to avoid classifying these mutants as defective members of their species when they exhibit the traits we consider naturally good in later populations for which they served as necessary precursors.

(ii) *The AN Solution*

The advocate of AN can, of course, bite the bullet and restrict the scope of Aristotelian categoricals to specific stages in the phylogenetic evolution of a species.⁶ Yet this would result in especially counterintuitive judgments of natural defect in cases of mutations that are clearly beneficial for the species in question, and it would inhibit the possibility of normative progress in a way that conflicts with even a common-sense understanding of species mutability. Hence, the more prudent response to the mutant problem is to appeal to the same resources that AN employs to reply to prior counterexamples based on evidence from evolutionary biology, that is, to reiterate that AN is immune to empirical defeaters because the natural history judgments that ground its content are informed

by *non-observational* claims. These claims are not exclusively based on the details of biological science, because they incorporate implicit postulations of life-form categories that impose coherence on our experiences of the natural world. When we specify the details of natural goodness for a species, according to AN, we do not merely list the statistically typical features of a population at some point in the phylogeny of a species. What philosophers like Thompson emphasize is that the content of AN life forms is primarily fixed by norms that are independent of these statistical features, because this content is not forced on us by nature—it is instead *presupposed* in our ascriptions of stable life cycles to the otherwise disordered information we collect from empirical observations. Consequently, mutants do not constitute defeating counterexamples for AN, because they can be assimilated within whatever stable species forms we presuppose in order to distinguish species from background patterns of natural phenomena.

(iii) *Empirical Complications*

This is not to say, however, that AN will be able to avoid complex questions that arise at the margins of species delineation. When, exactly, will novel mutations lead AN to decide that a population has become a new species? How, exactly, will AN determine whether a mutation is defective or not if the mutation contributes positively to survival and reproduction but does so in a way that is not typical given existing species norms? Questions like these are delicate for AN because its most plausible formulations incorporate non-observational insights with at least *some* information that we collect through empirical observations. As Daniel Groll and Micah Lott acknowledge:

Clearly, apprehending life-forms is not an *a priori*, armchair exercise; it requires actual fieldwork! [. . .] Of course our conception of the life-form ‘bobcat’ comes about from observ-

ing bobcats. It would be absurd to deny that. (2015, fn 23)

Similarly, Richard Kim (2018) notes that giving up the idea that empirical observations inform our knowledge of species life forms would be to leave out the naturalist basis of AN's normative foundation and leave the view vulnerable to the kind of wishful thinking that creates what Millgram calls the *Polyanna Problem*: the objection that AN relies on an overly optimistic conception of human nature by ignoring normatively unappealing aspects of our species (2009, p. 561). As Kim emphasizes, "One of the advantages of ethical naturalism over various forms of intuitionism is the avoidance of this kind of danger, by taking seriously the need to integrate theory with ground-level observations" (2018, p. 145). This incorporation of empirical details into the implicit presuppositions of norms for each life form allows AN to present a persuasive account of how we identify, and continuously revise, our classifications of species-typical traits, but this balance between non-observational presuppositions and contingent empirical details drives AN into some murky issues at the boundaries of species classifications.

For example, what should AN say about the first mutant vervet monkey who issued an alarm call that was specific to eagles rather than a general alarm call for all predators? In our prior case of meerkat alarm calls, an increase in the volume or the frequency of issuing a call when predators appear would be fairly easy for AN to describe as a mutation that is beneficial, rather than defective in virtue of being atypical, because this new mutation extends our *existing apprehension* of the meerkat life form. When we think of the kinds of vital activities that define the species, louder or more frequent alarm calls will only increase the odds that a meerkat will accomplish these activities in its own distinct way. A mutation in a vervet monkey that allows

it to issue predator-specific alarm calls, by contrast, is not just an improvement in the activities of that life form; it is a new and highly unique way for it to achieve its general aims of survival and reproduction (Price et. al., 2015). Clearly, it is not sufficient to describe *any* trait that improves reproductive fitness as a beneficial mutation, according to AN, for that would be to give up its Aristotelian foundation based on unique life cycles in favor of an instrumental, evolutionary focus on survival by any means necessary. As Groll and Lott articulate the point, "In general, then, what benefits a living thing seems to be whatever helps it to live well *as the kind of thing it is*—i.e. to do the things that make it good *as a member of its kind*, and enable it to live well *as that kind of organism*" (2015, p. 643, my emphasis). Thus, AN faces some tough decisions when it comes to classifying mutations that increase fitness in ways that cannot currently be described in ways that are characteristic of the life forms in question.

If the vervet example seems to rely on too subtle a distinction in alarm calls, we can turn to other possibilities. Consider vestigial traits that play a vital role in phylogenetic classification but are often detrimental to the survival and reproduction of existing species. If a mutation occurs that improves the fitness of an individual by eliminating a vestigial trait that is one of the distinctive features of its life form, can AN consider the mutation beneficial? As emus evolved into flightless birds, was it beneficial for each mutant to assign fewer resources to its wings during ontological development if this reallocation of resources made it adaptive in a new, distinctive way? If some group of mutant rabbits managed to gain a fitness advantage by relinquishing the indefinite benefits of an appendix for the competing advantage of not having to endure its risks, can AN consider this a beneficial modification given that possession of an appendix is a distinctive feature of the rabbit life form that is shared with only a

handful of other mammals? Moreover, similar questions arise for non-vestigial traits that are disadvantageous but highly characteristic of certain species. For example, if sexual selection pressures change so that male peacocks gain a fitness advantage by ceasing to drag around heavy, predator-attracting feathers (i.e. targets) on their rear ends, can AN recognize this as a *benefit* for the mutants who initiate the change in one of its flagship life forms?⁷

(iv) *The AN Solution Revisited*

As intriguing as these examples of mutants might be, they do not represent *defeating* counterexamples for AN, because in any case of mutation, no matter how atypical it may be, AN can incorporate the mutation into new standards of what is good for a species by revising the set of Aristotelian categoricals that we apply to the life form in question. It is true that AN will sometimes encounter difficult transition cases between stages of selection, but even if the view remains sensitive to the empirical details that make these cases complicated, there is no doubt for AN that our non-observational presuppositions of life form standards are not bound by these empirical details. Quite the contrary, the concept of a life form specifies the standards for what counts as good or defective for its members, and it is therefore *conceptually prior* to any details we acquire through empirical observations. As Groll and Lott claim, “if we want to evaluate an individual organism, we need to know what kind of organism we are dealing with” (2015, p. 640). Thus, no mutant could present defeating empirical details for AN, because these details can only be meaningfully interpreted as good or defective through the lens of its non-observational life form concept, and this concept is something that we determine rather than something imposed upon us by, say, the statistical averages of a species population.

In other words, we *author the content* of natural norms, according to AN, insofar as

we are responsible for the normative structure of the Aristotelian categoricals we require to make sense of our empirical observations, even if these empirical observations also play a vital role in the process of continuously testing and revising our life form concepts. For example, an emu with smaller than average wings is only a *defective* mutant until we update the norms we apply to our apprehension of its life form. Once we change our perspective and consider it a flightless bird, this very same characteristic becomes a *beneficial* allocation of limited resources. (Heers and Dial 2015) Similarly, classifying a mutant peacock with more or less colorful plumage will depend crucially on whether we consider this plumage to be paradigmatic of its life cycle. It is open to us to decide what level of plumage is presupposed to serve as the non-observational basis for the Aristotelian categoricals we apply to the species, and this remains true no matter what variation exists in its population or subsequent stages of phylogeny. The content of AN norms is, in this respect, safely immunized from the threat of mutant counterexamples.

Turning to our own species, it is important to acknowledge the profound way in which we apply pre-existing norms to our process of self-interpretation, according to AN, rather than merely deciphering these norms from the independent content of the natural world. This may seem like an unlikely methodology for a *naturalistic* theory of normativity, but supporters of AN proudly acknowledge that initial values are presupposed in the continuous interpretative task of deciphering norms that are emblematic of our unique life form (Hacker-Wright 2009a, p. 315; 2009b, p. 416; 2012, p. 20; Hursthouse 2012, pp. 174–77; Lott 2012a, pp. 372–74; 2012b, pp. 416–18). From this perspective, we need only relinquish whatever foundationalist expectations we might have of normatively inert natural facts determining what is prescriptive for our species to see that we are ultimately

responsible for the natural norms that govern our life form.

4. ALIENS

The second potential counterexample to AN is more straightforward: the view does not specify norms for sentient *non-humans* who exhibit capacities for practical reason like those that make moral norms authoritative for human agents. In other words, the content of AN prescriptions regarding human natural goodness and defect cannot be applied to agents like Martians, elves, thanagarians, or wookies. The previous AN reply to the mutant objection allows considerable breadth in the content of the norms we author when we engage in the apprehension of a life form, but moral norms are specific to the *human* life form, according to AN methodology, and therefore cannot directly apply to non-human cases. This leads to an odd result, given the AN focus on patterns of practical reason, because it seems counterintuitive to think that the view has nothing to offer when it comes to ascribing norms of justice, honesty, or beneficence to the kinds of hypothetical aliens who are presumed to be strikingly similar to humans.

(i) *The Threat of Kantian Assimilation*

The first possible reply to this objection is to claim that norms for humans will also apply to aliens in virtue of their status as creatures with capacities for practical reason. The objection is only convincing, this reply will point out, if the examples put forward are relevantly similar to humans rather than, say, the alien equivalents of mosquitos or turnips, and if the aliens at stake are rational agents then AN advocates may be tempted to assert that all rational agents, human or non-human, are subject to identical *sui generis* norms that govern any creatures who exhibit practical rationality as part of their characteristic life cycles.

Indeed, this is a plausible reply; however, it is not available to advocates of AN, because

opting for this reply surrenders the distinctive features of the view and tacitly substitutes it for a variety of Kantian ethics. As critics and advocates of AN alike have noted in the context of the *Pollyana Problem*, if Foot's view kicks away the ladder of a naturalist basis for human rationality and endorses norms of practical reason that are free of any contingency related to membership in a particular species, then the view is not merely related to Kantian ethics, as defenders of AN commonly acknowledge; the view just *is* Kantian ethics with a lofty but dispensable taxonomic backstory (Lott 2014, pp. 772–76; Woodcock 2015, pp. 25–29). Thus, AN cannot defend against aliens by retreating to an idealized conception of rational agency stripped of the contingent features of human nature, because doing so would succeed at the cost of no longer proposing a distinct position in the literature.⁸

(ii) *A Distinct AN Solution*

Fortunately for AN, it need not reply to alien counterexamples by appealing to norms grounded in a single, abstract conception of practical reason. Instead, AN advocates can argue that: (a) humans, along with each distinct alien species, are subject to their own unique norms of practical reason, and (b) we can reasonably speculate that there would be sufficient *overlap* between these separate sets of norms that we can expect basic similarities to exist for rational creatures with life cycles that resemble the human life cycle. Advocates of AN often explicitly assert that the view is species-relative (Foot 2004, p. 13; Lott 2012, p. 366; Hursthouse 2012, p. 178; Hacker-Wright 2013a, p. 85) and would therefore not apply to non-humans. Yet an emphasis on species-specific content for norms of practical reason need not entail that no basic similarities exist between creatures who coordinate in similar ways to survive and reproduce. Of course, some alien species may be too radically different from our own for

any overlap in their norms of practical reason to exist. Some alien species might even take the vicious forms of H.R. Giger nightmares.⁹ However, the possibility of overlap in the life cycles of humans and other species of creatures with capacities for practical reason is sufficient to dismiss the counterintuitive feel that alien counterexamples initially generate.

One might note at this point that the contingency for norms of human practical reason that distinguishes AN from Kantian ethics will only exacerbate the *Pollyana Problem*, for if the content of practical reason for human and alien species depends upon the slings and arrows of each species' unique developmental history, then it is difficult to view this content as a reliable source of normative guidance.¹⁰ I will not, however, pursue this problem for AN here. For now, my point is only that a persuasive AN reply to alien counterexamples—one that avoids the trap of surrendering the distinctive features of the view and opting for Kantian ethics—will allow for each species to exhibit practical reason in its own distinct, contingent ways that could lead to significant variation in normative content.

5. THE GREAT RED DRAGON

The third counterexample for AN will initially strike readers as peculiar, yet I will argue that the replies that AN gives to the first two counterexamples force the view into being unable to give a satisfactory reply to this peculiar third type of case for which an ethical theory ought to be able to offer a straightforward response. The case I have in mind is someone who is a human being, in an ordinary genetic sense, but who no longer considers him or herself to be a member of the human species. Specifically, consider the character Francis Dolarhyde from the Thomas Harris novel *Red Dragon* (1981). Dolarhyde is a serial murderer who preys on those he selects to bear witness to his transformation into the imposing figure from William Blake's

painting *The Great Red Dragon and the Woman Clothed in Sun*. Like many evil characters from the genre, he does not consider his actions morally wrong. Instead, he views his victims as having no more status than ants or slugs in the face of his great "Becoming." He explains to a victim, "But you see, I am not a man. I began as one but by the Grace of God and my own Will, I have become Other and More than a man" (218). Thus, Dolarhyde explicitly justifies his actions according to moral norms that follow from a process of self-interpretation in which he views himself as a separate life form with its own unique patterns of appropriate behavior.¹¹

What can a supporter of AN say about this case? I take it as given that Dolarhyde's acts ought to be unequivocally condemned, yet it is not clear that AN has the resources to describe his acts as morally wrong. To do so, for AN, would be to show that Dolarhyde exhibits a failure of practical reason according to norms that describe a characteristic life cycle for the species to which he belongs. However, as we have seen in our discussion of aliens, AN must allow norms of practical reason to be unique to each particular life form, so it must allow for the possibility that the norms governing members of the red dragon species ought to kill humans in order to reflect the incomparable greatness of the red dragon species' transformations. In the same way that it is characteristic for lions to kill hyenas for the sake of gaining competitive advantage as predators, it is conceivable that an ordinary part of the red dragon life cycle involves executing humans for the sake of receiving a befitting sense of awe from lesser beings.

(i) *Unavailable Reply #1: Empirical Details*

Of course, supporters of AN will at this point claim that the preceding counterexample is preposterous, since Dolarhyde is obviously a delusional member of the *human*

species. Yet the view runs into difficulty supporting this seemingly obvious claim. What evidence can it present to make the case that Dolarhyde is human? The easiest way to do so would be to emphasize Dolarhyde's genetic, morphological, and physiological features. This information would provide a straightforward means of establishing his membership in the human species; however, this is precisely the kind of information to which AN cannot appeal. As the prior discussion of mutants illustrated, advocates of AN emphatically distance themselves from empirical facts drawn from the biological sciences having decisive normative status. This is an essential feature of AN that immunizes the view from defeaters that arise if it is subjected to empirical scrutiny. The basis for establishing that an individual is bound by norms that are specific to its unique life cycle is *our own postulation* of Aristotelian categoricals that non-observationally demarcate the species according to values that we presuppose in the interpretative task of deciphering norms that are emblematic of our particular life form. Taken seriously, this methodology leaves AN vulnerable to the red dragon counterexample, because no principled constraints are placed on individuals who project a novel life cycle for themselves as part of their "natural" self-interpretation.

(ii) *Unavailable Reply #2:*

Rational Agency

By contrast, consider the reply to the red dragon counterexample available to Kantian ethics. Advocates of this view (and other versions of meta-normative constitutivism) need not worry about whether Dolarhyde is human or some other species. As long as he can exercise a capacity for practical reason, he is bound by norms that issue from his own agency. He cannot opt out of these norms because they are constitutive of his ability to employ practical reason to raise this skeptical challenge in the first place. To reject the

authority of the norms that govern rational agents is self-contradictory, according to this view, because this rejection both depends upon the exercise of rational agency and seeks to undermine the norms that make such agency possible. This Kantian constitutivist view takes various forms (Korsgaard 1996, 2009; O'Hagan 2004, 2014; Ferrero 2009), and it is not without its critics (FitzPatrick 2005; Enoch 2006; Tiffany 2012). Nevertheless, if membership in the moral community is determined according to one's capacity for practical reason, irrespective of one's species classification, then a principled reply is available to block the red dragon counterexample.

Unfortunately for AN, no similar constitutive bedrock exists to prevent agents from the decision to opt out of the norms the view presents as characteristic of human functioning. If an agent like Dolarhyde interprets himself as being something other than a member of the human species, no clear practical contradiction arises for his subsequent deliberations. Similarly, it is not incoherent for a person who routinely engages in convenient deception to see herself as a representative member of an Enoch-esque species called *shmumans*. It would certainly seem idiosyncratic, but if species identification genuinely results from the supposition of norms that filter our empirical observations into intelligible life cycles, then this person cannot be charged with committing any manifest error in practical reason.

Finally, note that when an agent like Dolarhyde, or our deceitful *shmuman*, asks, "Why be human?" there are two possible interpretations of this skeptical challenge. First, one might take the agent to be asking why she ought to identify with the norms that govern a species of which the agent recognizes herself as a member. According to this interpretation, the question can be rephrased as, "Why must I consider myself bound by norms that are characteristic of my own species?" According to the second interpretation, however, the initial "Why be human?" question literally

asks why one must consider oneself a member of that species. Advocates of AN have recognized the challenge associated with the first interpretation of the question (Lott 2014) and made efforts to meet this challenge employing methods resembling those of Kantian constitutivism (Hacker-Wright 2012).¹² I have doubts about the viability of these efforts, but for now my aim is only to emphasize that the challenge created by the red dragon counterexample refers to the *second* of these two interpretations. It is easy to overlook this difference, since it is unusual for the “Why be human?” question to ever be asked with the second interpretation in mind. Indeed, the red dragon objection to AN is a decidedly strange one. Nevertheless, the strange nature of the objection is not independently derived but is instead a reflection of what it reveals about the methodology AN employs to generate its normative content. It may seem advantageous for norms of species classification to follow from idealized suppositions that are conceptually prior to the contingent details of empirical biology and the trappings of objections based on mutants or the Pollyanna problem, but this apparent advantage leads to an exceedingly odd inability to set firm limits on species classification and a consequent vulnerability to the objection represented by the Great Red Dragon.

6. THE MEMBERSHIP OBJECTION

The combination of the three counterexamples presented in this paper generates the problem that I describe as a *membership* objection. The objection is that AN invites a trilemma when it is pressed to consider agents at the edges of membership in its species categories. First, if AN relies on straightforward empirical criteria for determining what it takes to be the characteristic features of the members of its species, then this leads to counterintuitive results faced with the prospect of mutants who exhibit atypical features that may become archetypal

at some future stage of phylogeny. Second, if AN accounts for aliens with features similar to those of humans by appealing to abstract rational agency, stripped of the contingent details associated with each species’ unique developmental history, then AN collapses into a type of Kantian ethics with a dispensable naturalistic anthropology. I believe AN can successfully avoid these first two horns of the trilemma; however, to do so the view must emphasize aspects of AN that lead to a third, counterintuitive conception of species membership. If (a) we are ultimately responsible for the normative patterns that specify Aristotelian life cycles and make sense of subsequent empirical observations, and (b) if the content of these life cycles is species-specific so that a non-human agent exhibiting practical rationality might be governed by a significantly different set of natural norms, then AN must accept the fact that a human individual, who could be a new mutant for whom new normative standards apply, can engage in a process of self-interpretation that gives rise to norms that diverge radically from those we normally associate with our common-sense understanding of the human species.

What the membership objection reveals is that the persuasiveness of AN depends upon an equivocation in its appeal to facts about the human species. To avoid counterexamples like mutants (and various ways of pressing the *Pollyanna* problem), AN must emphasize that it relies on an interpretation of natural normativity derived from Michael Thompson according to which we ultimately author the non-observational facts of our life form. However, to avoid opening up AN to counterexamples like the Great Red Dragon, the view must tacitly rely on an ordinary biological conception of the human species so that fixed normative content and inescapable species membership can be presumed. Once this equivocation is made explicit, it is clear that AN cannot have it both ways.

Either the view commits itself to a unified view of normativity for the human species and accepts the unwanted baggage that comes with determinate content and non-optional membership, or else the view allows for freedom of self-interpretation in its normativity that allows for self-proclaimed mutants like Francis Dolarhyde.

Advocates of AN will surely find this result absurd. Yet this speaks, I think, to the oddly intransigent nature of the debate between AN and its critics. Advocates of AN, for their part, consider each objection to AN to depend upon some avoidable misinterpretation of the view, and if presented with the Red Dragon objection one can expect them to claim that it rests on the false presumption that AN allows agents to engage in a process of self-interpretation that is unconstrained by basic facts about the human species. As we have seen, advocates like Groll and Lott emphasize that AN is not committed to Aristotelian

life-forms that we apprehend via an armchair exercise that would allow Dolarhyde to invent his own “natural” norms. They will insist that the norms that specify Aristotelian life forms are discovered via an *empirical* process that results in determinate facts about our species that are not open to arbitrary interpretation by agents like Dolarhyde.¹³

Critics of AN, for their part, find it frustrating that clarifications of the view often seem to work in ways that contradict one another. Thus, I hope to break the apparent stalemate by proposing an objection to AN that demands *consistency across its various replies* to individual objections. As convincing as each AN reply to a particular objection might appear in isolation, the *membership* objection is designed to capture the fact that each new reply leads to a further objection, making the overall experience of arguing against AN feel like a continuous game of whack-a-mole represented by Figure 1.

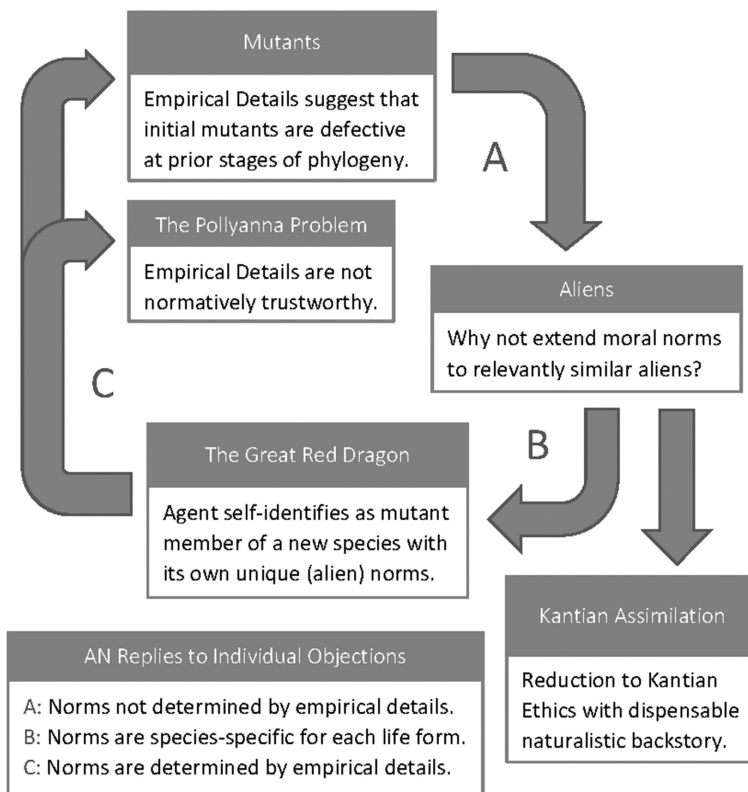


Figure 1. The Membership Objection

When the various replies from AN are represented simultaneously, it becomes apparent that the replies that seem so decisive when viewed in isolation tend to exacerbate other objections to the view. Thus, the *membership* objection is designed to undermine the global plausibility of AN by establishing a burden of proof against it being able to provide a consistent interpretation of itself that grounds all of its various replies against defeating objections. It may be possible for AN to offer a justification for all of these replies simultaneously without resorting to some sort of promissory note regarding species-specific norms that are empirically discoverable and normatively trustworthy, but the membership objection suggests that this is a task that will require further arguments from AN than have been offered to date.

7. EPILOGUE

In closing, it is worth noting that it is not inconceivable for advocates of AN to opt for the third horn of the trilemma above and *accept* that AN is unable to condemn the behavior of an agent as deranged as Francis Dolarhyde. It would not be the first theory to fail to convince delusional psychopaths to behave morally, so what is so damaging about surrendering to the red dragon counterexample? First, it is important to note that the issue is not whether AN can convince an agent like Dolarhyde to behave morally. No theory can change the minds of agents who refuse to listen to reason. The issue, rather, is whether the theory provides an adequate justification in principle for such agents. The problem for AN is that its normative judgments are contingent on species membership in a problematic way given its methodology. Still, the option is open to advocates of AN to consider the imperatives issued by the theory as *hypothetically* dependent on agents' identification

with the human species. The theory would be unable to convince red dragons and other less extreme examples of people crafty enough to interpret themselves as, say, *shmunans*, but if other prominent meta-ethical proposals issue no more than hypothetical imperatives (e.g. Railton 1986), then supporters of AN can certainly do the same.

I suspect, however, that most supporters of AN will, like Foot, reject this option as less than satisfactory.¹⁴ Foot opens *Natural Goodness* with the explicit aim of defending a view of practical rationality that surpasses her own prior efforts (1972) to give force to our normative obligations (2001, pp. 9–11). Moreover, she describes a prominent motivation of her work as the aim of being able to provide substantive moral assertions to corrupt individuals like Nazis who refuse to comply with norms as basic as the prohibition on killing innocent persons (2003). In this context, the less substantive options to which Foot compares her own were the versions of non-cognitivism that dominated meta-ethics for the first half of the twentieth century. Yet it is reasonable to presume that Foot would be similarly underwhelmed by a conception of natural normativity that issued only hypothetical imperatives to its agents by permitting the option of opting out of Aristotelian categoricals for the human species if one interprets oneself to be a member of some other life form. Thus, opting for the third horn of the membership objection trilemma is a possible but almost certainly unappealing option for AN. This, I conclude, leaves AN in a difficult position, since the first two horns of the trilemma look equally unappealing, if not considerably worse, when it comes to inviting counterexamples that undermine the view's normative appeal.

University of Victoria

NOTES

I would like to thank two anonymous referees from this journal for their very thoughtful comments on previous drafts of this paper.

1. One also encounters less explicit defenses of Foot that are nonetheless sympathetic to her naturalistic approach (e.g. Moore 2004; Sandler 2005; Thompson 2007). Most prominently, Rosalind Hursthouse (1999, 2012) defends a form of Aristotelian Naturalism inspired by Foot but with its own distinct features.
2. For a helpful overview of these classification issues in the context of human nature, see (Kronfeldner, Roughley, and Toepfer 2014).
3. See also Andreou (2006, pp. 68–72), (Millum 2006, pp. 211–12) and Glackin (forthcoming, s2).
4. This is not to say that this AN reply is uncontroversial. I argue elsewhere (2015) that this reply risks introducing a level of indeterminacy into AN that drains the view of informative content.
5. For the sake of simplicity I set aside mechanisms of evolutionary change other than graduated versions of natural selection. This is a controversial assumption in the philosophy of biology, but alternative mechanisms like those famously emphasized by Richard Lewontin and Stephen Jay Gould only add to the philosophical challenge at stake for AN. For example, if random drift or mass extinction via meteor impact is responsible for causing changes to the phenotype of a species (Gould and Lewontin 1979; Gould 2002), then it would be that much more idiosyncratic for AN to refer to a trait as defective at t1 when the same trait will be emblematic of the species at t2.
6. Foot appears to endorse this strategy (2001, p. 29); however, it is not clear that she had the opportunity to consider the prospect of mutants when she committed herself to this time-indexed view of species.
7. More fanciful examples are also available to philosophers with active imaginations. What if, in some unlikely case of punctuated equilibrium, a precursor mutant develops a trait like telekinesis, retractable claws, or the capacity to control weather patterns? Can a trait like one of these, even if it later reaches fixation, be described as beneficial by AN when it so radically diverges from what is characteristic of a human living well *as a member of its kind*?
8. For helpful comparisons of AN to Kantian ethics, see (Lebar 2008) and (Korsgaard 2011, pp. 382–86).
9. The point here is reminiscent of Gilbert Harman's influential discussion of aliens who might not share any of our standards of practical reason (1975, p. 5). Yet Harman's more specific point about the obligations of moral judgment being contingent on the *motivational attitudes* of agents need not be endorsed for the more general point to hold regarding moral relativism between human and alien species.
10. Milgram's invocation of the *Wason Selection Task* is particularly insightful here (2009, p. 563).
11. Friedrich Nietzsche's *Übermensch* (2006) springs to mind here as a similar example; however, I will use Dolarhyde to avoid the complex interpretive issues that arise for the question of whether Nietzsche presents the *Übermensch* as an ideal for which all humans ought to strive or as a separate being who is described as transcending humanity.
12. An influential articulation of this challenge is found in (McDowell 1995) and further discussed in (Toner 2008).
13. Advocates of AN will also emphasize that we, as an overall human community, employ an epistemic division of labor between specialists and non-specialists for the task of discovering the empirical content of species-specific norms. We turn to trained biologists to uncover the characteristic facts about, say,

the alpine pika's life cycle, and we analogously have no reason to think that an individual agent like Dolarhyde is the right kind of specialist to be competent to determine the norms that apply to the life form that he is. I thank an anonymous referee for noting this point.

14. A notable exception here might be (Hendley 2015). By contrast, Stephen Finlay summarily dismisses naturalistic theories of normativity based on human function because of their inability to capture the specifically categorical sense of "ought" (2009, p. 332).

REFERENCES

- Andreou, Chrisoula. 2006. "Getting on in a Varied World," *Social Theory & Practice*, vol. 32, no. 1, pp. 61–73.
- Brown, Stephen R. 2004. "Naturalized Virtue Ethics and the Epistemological Gap," *Journal of Moral Philosophy*, vol. 1, no. 2, pp. 197–209.
- Copp, David, and David Sobel. 2004. "Morality and Virtue: An Assessment of Some Recent Work in Virtue Ethics," *Ethics*, vol. 114, no. 3, pp. 514–54.
- Enoch, David. 2006. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action," *The Philosophical Review*, vol. 115, no. 2, pp. 169–98.
- Ereshefsky, Marc. 2001. *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy* (Cambridge: Cambridge University Press).
- . 2009. "Darwin's Solution to the Species Problem," *Synthese*, vol. 175, no. 3, pp. 405–25.
- . 2010. "What's Wrong with the New Biological Essentialism," *Philosophy of Science*, vol. 77, no. 5, pp. 674–85.
- Ferrero, Luca. 2009. "Constitutivism and the Inescapability of Agency," in *Oxford Studies in Metaethics*, Vol. 4, ed. Russ Shafer-Landau (Oxford: Oxford University Press), pp. 303–33.
- Finlay, Stephen. 2009. "Oughts and Ends," *Philosophical Studies*, vol. 143, no. 3, pp. 315–40.
- Fitzpatrick, William. 2000. *Teleology and the Norms of Nature* (New York: Garland Publishing).
- . 2005. "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity," *Ethics*, vol. 115, no. 4, pp. 651–91.
- Foot, Philippa. 1972. "Morality as a System of Hypothetical Imperatives," *The Philosophical Review*, vol. 81, no. 3, pp. 305–16.
- . 2001. *Natural Goodness* (Oxford: Oxford University Press).
- . 2003. "The Grammar of Goodness," *The Harvard Review of Philosophy*, vol. 11, no. 1, pp. 32–44.
- . 2004. "Rationality and Goodness," *Royal Institute of Philosophy Supplements*, vol. 54, pp. 1–13.
- Glackin, Shane N. 2015. "Three Aristotelian Accounts of Disease and Disability," *Journal of Applied Philosophy*, vol. 33, no. 3, pp. 311–26.
- Gould, Stephen Jay. 2002. *The Structure of Evolutionary Theory* (Cambridge, Mass.: Harvard University Press).
- Gould, Stephen Jay, and R. C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," *Proceedings of the Royal Society of London*, vol. 205, no. 1161, pp. 581–98.
- Groll, Daniel and Micah Lott. 2015. "Is There a Role for 'Human Nature' in Debates About Human Enhancement?" *Philosophy*, vol. 90, no. 4, pp. 623–51.
- Hacker-Wright, John. 2009a. "What Is Natural About Foot's Ethical Naturalism?" *Ratio*, vol. 22, no. 3, pp. 308–21.
- . 2009b. "Human Nature, Personhood, and Ethical Naturalism," *Philosophy*, vol. 84, no. 3, pp. 413–27.
- . 2012. "Ethical Naturalism and the Constitution of Agency," *The Journal of Value Inquiry*, vol. 46, no. 1, pp. 13–23.

- . 2013a. “Human Nature, Virtue and Rationality,” in *Aristotelian Ethics in Contemporary Perspective*, ed. Julia Peters (New York: Routledge), pp. 83–96.
- . 2013b. *Philippa Foot’s Moral Thought* (London: Bloomsbury).
- Harman, Gilbert. 1975. “Moral Relativism Defended,” *The Philosophical Review*, vol. 84, no. 1, pp. 3–22.
- Harris, Thomas. 1981. *Red Dragon* (New York: Dell Publishing).
- Heers, A. M. and Dial, K. P. 2015. “Wings Versus Legs in the Avian Bauplan: Development and Evolution of Alternative Locomotor Strategies,” *Evolution*, vol. 69, no. 2, pp. 305–20.
- Hendley, Steven. 2009. “Reassuring Ourselves of the Reality of Ethical Reasons: What McDowell Should Take from Foot’s Ethical Naturalism,” *Dialogue*, vol. 48, no. 3, pp. 513–37.
- . 2015. “Moral Reasoning as Naturally Good: A Qualified Defense of Foot’s Conception of Practical Rationality,” *Southern Journal of Philosophy*, vol. 53, no. 4, pp. 427–49.
- Hollén, Linda L., and A. N. Radford. 2009. “The Development of Alarm Call Behaviour in Mammals and Birds,” *Animal Behaviour*, vol. 78, no. 4, pp. 791–800.
- Hull, David L. 1965a. “The Effect of Essentialism on Taxonomy—Two Thousand Years of Stasis (I),” *The British Journal for the Philosophy of Science*, vol. 15, no. 60, pp. 314–26.
- Hull, David L. 1965b. “The Effect of Essentialism on Taxonomy—Two Thousand Years of Stasis (II),” *The British Journal for the Philosophy of Science*, vol. 16, no. 61, pp. 1–18.
- Hursthouse, Rosalind. 1999. *On Virtue Ethics* (Oxford: Oxford University Press).
- . 2012. “Human Nature and Aristotelian Virtue Ethics,” *Royal Institute of Philosophy Supplements*, vol. 70, pp. 169–88.
- Kim, Richard T. 2018. “Human Nature and Moral Sprouts: Mencius on the Pollyanna Problem,” *Pacific Philosophical Quarterly*, vol. 99, pp. 140–62.
- Kitcher, Philip. 1984. “Species,” *Philosophy of Science*, vol. 51, no. 2, pp. 308–33.
- Korsgaard, Christine M. 1996. *The Sources of Normativity* (New York: Cambridge University Press).
- . 2009. *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press).
- . 2011. “Natural Goodness, Rightness, and the Intersubjectivity of Reason,” *Metaphilosophy*, vol. 42, no. 4, pp. 381–94.
- Kronfeldner, Maria, Roughley, N. and Toepfer, G. 2014. “Recent Work on Human Nature: Beyond Traditional Essences,” *Philosophy Compass*, vol. 9, no. 9, pp. 642–52.
- LeBar, Mark. 2008. “Aristotelian Constructivism,” *Social Philosophy and Policy*, vol. 25, no. 1, pp. 182–213.
- Lewens, Tim. 2010. “Foot Note,” *Analysis*, vol. 70, no. 3, pp. 468–73.
- Lott, Micah. 2012a. “Have Elephant Seals Refuted Aristotle? Nature, Function, and Moral Goodness,” *Journal of Moral Philosophy*, vol. 9, no. 3, pp. 353–75.
- . 2012b. “Moral Virtue as Knowledge of Human Form,” *Social Theory and Practice*, vol. 38, no. 3, pp. 407–31.
- . 2014. “Why Be a Good Human Being? Natural Goodness, Reason, and the Authority of Human Nature,” *Philosophia*, vol. 42, no. 3, pp. 761–77.
- MacIntyre, Alasdair. 2002. “Virtues in Foot and Geach,” *The Philosophical Quarterly*, vol. 52, no. 209, pp. 621–31.
- McDowell, John. 1995. “Two Sorts of Naturalism,” in *Virtues and Reasons*, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn (Oxford: Clarendon Press), pp. 149–79.
- Millgram, Elijah. 2009. “Life and Action,” *Analysis*, vol. 69, no. 3, pp. 557–64.
- Millum, Joseph. 2006. “Natural Goodness and Natural Evil,” *Ratio*, vol. 19, no. 2, pp. 199–213.
- Moore, Adam D. 2004. “Values, Objectivity, and Relationalism,” *The Journal of Value Inquiry*, vol. 38, no. 1, pp. 75–90.
- Nietzsche, Friedrich. 2006. *Thus Spoke Zarathustra* (Cambridge: Cambridge University Press).
- O’Hagan, Emer. 2004. “Practical Identity and the Constitution of Agency,” *The Journal of Value Inquiry*, vol. 38, no. 1, pp. 49–59.

- . 2014. “Shmagents, Realism and Constitutivism About Rational Norms,” *The Journal of Value Inquiry*, vol. 48, no. 1, pp. 17–31.
- Okasha, Samir. 2002. “Darwinian Metaphysics: Species and the Question of Essentialism,” *Synthese*, vol. 131, no. 2, pp. 191–213.
- Price, Tabitha, Philip Wadewitz, Dorothy Cheney, Robert Seyfarth, Kurt Hammerschmidt, and Julia Fischer. 2015. “Vervets Revisited: A Quantitative Analysis of Alarm Call Structure and Context Specificity,” *Scientific Reports*, vol. 5, no. 13220.
- Railton, Peter. 1986. “Moral Realism,” *The Philosophical Review*, vol. 95, no. 2, pp. 163–207.
- Sandler, Ronald. 2005. “What Makes a Character Trait a Virtue?” *The Journal of Value Inquiry*, vol. 39, no. 3, pp. 383–97.
- Slote, Michael. 2003. “Review: Natural Goodness,” *Mind*, vol. 112, no. 445, pp. 130–39.
- Sober, Elliott. 1980. “Evolution, Population Thinking, and Essentialism,” *Philosophy of Science*, vol. 47, no. 3, pp. 350–83.
- Teichmann, Roger. 2011. *Nature, Reason, and the Good Life: Ethics for Human Beings* (Oxford University Press).
- Thompson, Allen. 2007. “Reconciling Themes in Neo-Aristotelian Meta-Ethics,” *The Journal of Value Inquiry*, vol. 41, no. 2, pp. 245–63.
- Thompson, Michael. 1995. “The Representation of Life,” in *Virtues and Reasons: Philippa Foot and Moral Theory*, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn (Oxford: Clarendon Press), pp. 247–96.
- . 2003. “Three Degrees of Natural Goodness,” *Iride*. Originally published in Italian. Available at: <http://www.pitt.edu/~mthomps/three.pdf>.
- . 2004. “Apprehending Human Form,” in *Modern Moral Philosophy*, ed. Anthony O’Hear (Cambridge: Cambridge University Press), pp. 47–74.
- . 2008. *Life and Action: Elementary Structures of Practice and Practical Thought* (Cambridge, Mass.: Harvard University Press).
- Tiffany, Evan. 2012. “Why Be an Agent?” *Australasian Journal of Philosophy*, vol. 90, no. 2, pp. 223–33.
- Toner, Christopher. 2008. “Sorts of Naturalism: Requirements for a Successful Theory,” *Metaphilosophy*, vol. 39, no. 2, pp. 220–50.
- Wilkins, John S. 2009. *Species: A History of the Idea* (Berkeley: University of California Press).
- Woodcock, Scott. 2006. “Philippa Foot’s Virtue Ethics Has an Achilles’ Heel,” *Dialogue*, vol. 45, no. 3, pp. 445–68.
- . 2015. “Neo-Aristotelian Naturalism and the Indeterminacy Objection,” *International Journal of Philosophical Studies*, vol. 23, no. 1, pp. 20–41.