University of St Andrews
The Philosophy Society
APORIA

Vol. X

ABOUT

ACKNOWLEDGEMENTS

FROM THE EDITOR

Aporia was first published in February 2007. After five years, we have become the only student-run philosophy journal published in the United Kingdom that solicits international submissions. We are a forum for informed discussion on topics relevant to both philosophy professionals and the public alike. As always, article selection was guided by five desiderata: (1) urgency, (2) originality, (3) intelligence, (4) creativity, and (5) precision.

The purpose of this journal is to exchange philosophical ideas, as well as to highlight the importance of philosophy. Hopefully, this issue encourages others to study this magnificent subject.

In dedication to the red gowns of Saint Andrews.

STATISTICS, IN BRIEF: 34 submissions were received from 11 institutions, covering 7 countries. As such, the acceptance rate was 14.7%. All submissions were read blindly, and refereed anonymously.

CONTENTS

1 | GENDER DISINTEGRATION: PERFORMANCE, CONTEXT AND THE BODY

Tano Posteraro,[1] *McMaster University*

In her *Second Sex*, Simone de Beauvoir writes that "one is not born, but rather becomes, a woman" (1989: 267). With these words, de Beauvoir famously distinguished sex from gender. One may be born with a particular biological makeup, but it is socialisation that forms and regulates identity. In the following paragraphs, I will first explore the theory by which gender emerges as a social production, before considering how such an understanding leads to the problem of gender disintegration. I will then entertain Linda Alcoff's hermeneutics as a response to this problematic. By understanding perception to be shaped by the body, relative to context, I will conclude that Alcoff's hermeneutics is able to resolve the problem of gender disintegration.

Gender performativity is a concept attributed to contemporary philosopher Judith Butler, who defines gender as the effect, rather than the cause, of our actions and performances (1990: preface xv). For Butler, the female gender is the result of habitual action that produces the appearance of an anterior femininity. The

---

1 Tano Posteraro is a first year Masters student at *McMaster University*. His interests range from Nietzsche to phenomenology, feminist thought to the more recent writings of Gilles Deleuze. This paper, on the question of a feminist hermeneutic, was written in response to the gender-disintegration problematic.

performance that is itself responsible for the creation of gender is apprehended, instead, as its effect. Consider the observations of ethnomethodologist Harold Garfinkel. Garfinkel seeks to suspend indefinitely the belief in the reality of "normal women", and consider instead how individuals *produce* that reality for others (1967: 122).[2] In other words, he suspends sex as a given, focusing instead on how genders are produced. With Butler's performative theory in mind, Garfinkel turns to Agnes, the female-bodied, penis-endowed nineteen year-old who identifies as a woman. Because of her penis, Agnes must consciously present herself in a way that ensures no one will question the reality of her femalehood. We learn that Agnes produces the reality of her gender for those around her by considering how to act in any given social context: which interests to take up, in which fashion to speak, how to behave, walk, sit, eat, converse, and so on. In consciously deliberating her actions, Agnes is, Garfinkel concludes, performing her gender (1967: 119). Through this performance, she constructs the reality of her femininity for those around her. Performative theory maintains that there is no objective female gender. Rather, the female gender emerges out of a performance of particular actions. Agnes seems to perform these actions as if reading them from a script.

For performative theory, gender is the effect of the actions of an un-gendered subject. But, not all of a subject's actions may help constitute her gender. If we consider the endless multiplicity of a given subject's performances, (the way a person acts at work may differ from the way she acts at home or school) coupled with the observation that while some actions may constitute the production of one's gender, others may figure in the production of different identities entirely. We find ourselves in the midst of the problem of gender disintegration. In short: to say that gender has disintegrated is to say that it no longer retains any significance. Gender is one identity-constituting factor among several; as such, its significance is either wholly illusory or too difficult to separate out from other factors to remain meaningful. For example: is Agnes's love for cookbooks part of her gender performance or does it arise out of the multiple other identities

---

2 A gesture borrowed from Husserlian phenomenology: to properly investigate the contents of consciousness, Husserl first bracketed belief in the external world.

she may assume at a given time? At work, Agnes might like to converse about politics. Is she doing so because of an academic interest in political science, a capitalistic interest in economics, or a friendly inclination to engage others in compelling debate? It becomes increasingly difficult to apprehend and classify performance when we consider the intersection of gender (performativity) with other identity-constituting factors such as race, class, and upbringing. If all that stands between men and women is habitual action in accordance with a gender that exists only as an effect of those actions themselves, and the significance of any particular set of actions is not easily recognisable, then, what kind of foundation can feminism claim for itself? Gender must, therefore, be regrounded in the subject prior to performance if it is to retain its significance for feminism. Let us now turn to Linda Alcoff and her conception of a bodily hermeneutic.

Alcoff's understanding of hermeneutics picks up on Gadamer's concept of understanding through horizons, a theory in which a person's horizons are constituted by her or his culture, history, upbringing and social role (1975: 301). Alcoff defines a person's horizons as "the framing assumptions we bring with us to perception and understanding, the congealed experiences that become premises by which we strive to make sense of the world" (2006: 95). We develop these "framing assumptions" through socialisation, picking up on the language and presuppositions of those who socialise us. It is through this socialisation that we form our horizons, and thereby our horizonal understanding of the world around us. To better understand how a person's horizons affect their perceptions and understanding, Alcoff offers the examples of a queen and her servant (2006: 96). The queen, upon looking into her castle's dining area, sees a long elegant table suitable for entertaining many guests and accommodating many feasts. Her servant, however, looks upon the same table and sees only a nightmare to clean, an offensively large receptacle for dirt and dust. The queen and the servant interpret the same object (the same dining table) differently; they perceive and understand it through different horizons, framed by different assumptions.

Their antithetical perceptions, claims Alcoff, are constituted by the way they were socialised. Raised in royalty, the Queen knows only of feasts and parties, whereas the servant, raised in poverty, knows only of housework and practicality. Thus, their

differing horizons lead them to differing perceptions, a concept which Gadamer proposes serves to highlight the difference in understanding between individuals.

To this traditional hermeneutical understanding, which limits a person's horizons to their culture, history, upbringing and social roles, Alcoff adds the body as a horizonal constituting agent and concept (2006: 102). It is through the addition of the body that I posit Alcoff as a response to Butler. If gender emerges out of social performance, and Gadamer's hermeneutics posit the constitution of horizons by strictly social means, then taken apart from socialisation, gender is still rendered meaningless. What we need, then, is a more fundamental understanding of gender if we are to keep it from collapsing in its intersection with other horizonal-constituting agents (e.g., race, class, upbringing, etc.). In bringing the *body* to hermeneutics, Alcoff gives us exactly this: a primordial conception of gender (that is, one that thinks gender anterior to socialisation). For Alcoff, the way we move and function, bodily speaking, helps constitute our horizonal interpretation of the world around us. We apply terms derived from bodily experience to the things we encounter. For example, we say that the stocks are falling, like a body through space, when they are fundamentally only decreasing incrementally in number or percentage. Contrarily, when they are rising in value, we speak as if they have physically jumped. We are not merely neutral subjects faced with a social world, but rather *embodied* subjects; thus, "bodily experiences establish horizons just as traditions and cultures do" (2006: 102).

Of course, bodily experiences include concepts more complex than merely jumping or falling. The female body includes the experience of lactation and pregnancy, whereas the male body does not. Consequently, Alcoff believes these inherent differences reflect a "perceptual orientation and a conceptual mapping that determines value, relevance, and imaginable possibilities"(2006: 107). Just as the servant and Queen, possessing different horizons, are unable to perceive the same dining table, a girl learning of her ability to bear children will be able to imagine different possibilities, dangers and threats than a boy learning he will never have to. Our very embodiment, then, helps constitute the horizons we use to understand the world: nursing, childbirth and rape form the horizons of a female as a direct result of her body.

It is in here where we can locate gender. However, Alcoff does not deny that gender intersects with other identity-constituting factors such as race, and class. So, she allows that the gender of different women may constitute their horizons in different ways, just as one woman may interpret the possibility of childbearing as the highest calling of the human condition, whereas another may interpret it as an oppressive agent serving only to trap women in imminent existence.[3] Nonetheless, Alcoff insists that the possibility itself has a bearing on a woman's horizons, even given its intersection with other factors. Indeed, thought in this way, horizons are gendered and therefore able to provide a basis for solidarity among women.

To understand horizons better, Gadamer proposes that we can look more closely at our reading of the texts that orient our horizonal understandings (1975: 294). The constitution of one's horizons is a circular process: our understanding of who we are depends on the contexts in which we live and act, and, our understanding of how we live and act in those contexts depends, in turn, upon our understanding of who we are. Yet, we live and act in many contexts: in a family, as part of a society, in a career, for a cause, and so on. Thus, we can tell different stories of who we are depending on which context we choose to speak from. Gender, as a way of understanding who we are, is bound to context; it is never acontextual. This is the fundamental contribution of Alcoffian hermeneutics to the problem of gender disintegration. The way we understand gender is inextricable from the context through which we choose to view it. Therefore, there can be no gender *as such*; no objective womanhood or femininity. But, given the context of traditional familial roles, we can point to a more definite understanding of the significance of one's gender.*For example, there are certain roles fulfilled by a sister in a traditional Christian household, by virtue of her being female (that is, a sister).* Given the context of walking alone, late at night, a subject's body constitutes the way they construct the experience. That is, a woman with a vagina is more likely to fear sexual assault, and will therefore, meet the walk home with anxiety (her body determines her social experience). Therefore, context be-

---

3 Simone de Beauvoir theorised that traditional female gender roles such as housework and child-raising doomed women to an existence devoid of transcendence, and therefore, freedom: *imminence*.

comes a lens through which we can approach gender, and as contexts change, (as lenses are replaced) gender begins to change and appear differently. Likewise, just as with a camera without a lens, there is no picture, so without a context, there is no gender.

Further, the aforemantioned is complicated by the fact that context itself is also subject to interpretation. Our understanding of a female as a woman, then, will depend upon our understanding of the context in which femalehood is a part. Given the context of sexual reproduction, for example, the way we understand a female in terms of her sexual reproductive capacity is dependent on how we understand sexual reproduction itself. For Judith Butler, sexual reproduction is a "compulsory heterosexuality".[4] A female is a female in terms of sexual reproduction, insofar as she is capable of contributing an egg, and incubating a fetus. Moreover, as was outlined above, textual understanding is a circular affair, and so, if we can know a female as a child-bearer in terms of sexual reproduction, we can only know sexual reproduction in terms of its parts; namely: the female child-bearer and the male sperm-provider. Thus, just as the child-bearer is inextricable from the reproductive context, so the reproductive context is inextricable from the child-bearer. In short, sex becomes meaningless outside of such a relation. So, not only is such a gender-understanding derived from sex, but such a sex-understanding is, too, derived from gender. Consequently, taken apart from context, gender disintegrates.

Does this mean we have come full circle, have we undermined our own enterprise, refuted our own position? No, for we have brought the term *context* to the formula of disintegration. That gender becomes meaningless outside of context need not be negative. Consider the notion that apart from context, *everything* disintegrates; or more properly, that we can only come to know anything relative to its respective context. If this proposition can be coherently entertained, then gender becomes just as stable as anything else. This is not as radical as it sounds, and is not without its benefits. In terms of a medical context, we should consider people as patients and not as females or males, just as we would think it absurd to consider them either as cell-phone enthusiasts

_____

4 This means to think in terms of those body parts significant in the act of sexual reproduction.

or techno-luddites when ordering an organ-donor list.[5] Further, the introduction of context to gender relieves feminism of its tendency to homogenise women as inherently oppressed: as victims, not of history, but of nature itself.[6] By contextualising gender, we can do away with such essentialism, for even the most staunchly radical feminist will grant that in the context of finding work as a baby-sitter, women are not inhibited by their oppression, they are not oppressed at all. This crucial insight allows us to consolidate and refocus the feminist project. By addressing the particular contexts within which female oppression operates, we can begin to work more productively and effectively as feminists.

In short, if Alcoffian hermeneutics is to curb the threat of gender disintegration, we must hold that context *grounds* meaning.[7] This is to say that gender means different things in different contexts; apart from context, it means nothing at all. Thus, to give meaning to gender, we must first situate it. Far from demolishing gender, we have instead only contextualised it. Or perhaps it was a contextual notion to begin with. Perhaps performative theory only plucked gender from context, thereby emptying it of its significance. By reading Alcoff against Butler, perhaps we have only returned gender to its original home, granted it back its original meaning. Perhaps there truly is "nothing outside the text" (1976: 158-9).[8]

BIBLIOGRAPHY

[1] Alcoff, Linda. 2006. *Visible identities: Race, Gender, and the Self*. Oxford University Press: London.

---

5 This is to say that identity itself is always-already contextual. For example: in a familial context, I might be a disobedient son whose every move incites suspicion, while in an academic context, I might be a worthy student whose every comment demands attention. Thus, to speak of my disobedience *already implies* some specific context.

6 In other words, that women are victims not merely *historically*, but rather *naturally*.

7 Refer to the camera-lens metaphor presented above.

8 This is to say, of course, that there is nothing outside of discursive context, concerning gender. If it is to be present at all, it must be so within a context. Outside of that "text", there is *nothing*.

[2] Beauvoir, Simone de. 1989. *The Second Sex*. Vintage: New York.

[3] Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. Routledge Press: New York.

[4] Derrida, Jacques. 1976. *Of Grammatology*. Johns Hopkins University Press: Baltimore, Md.

[5] Gadamer, Hans. 1975. *Truth and Method*. Seabury Press: New York.

[6] Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Prentice-Hall, Inc.: Englewood Cliffs, NJ.

2 | AGATHON'S UNFULFILLED POTENTIAL

A Study of Agathon's contribution to the Symposium and its critique of Athenian Education

Sam Hege,[1] *University of Edinburgh*

1. INTRODUCTION

The unique construction of the *Symposium* has produced many different approaches to its characters and their roles. There have been countless interpretations of Socrates, Diotima, Alcibiades and Aristophanes's speeches and roles in the dialogue.[2] Scholars have also examined some of the less prominent speeches and characters. For example, David Konstan developed a powerful analysis of Eryximachus's speech that demonstrates its "intellectual rigor" and "the logic of the discourse" (Konstan, 1982: 44). Further, he shows how Eryximachus makes a legitimate move from "medical theory to some rather grandiose propositions about the cosmos and the gods" (Konstan, 1982: 44). However, even within these investigations, Agathon has for the most part, simply been set to the side, and consequently, interpretations have struggled to gain a coherent understanding of the dialogue.[3] At first glance, this inclination to dismiss Agathon cannot be faulted. He appears to be a rather uncontroversial and a useless character. His speech, for example, seems to be merely a

---

1 Sam Hege is currently reading for a Master of Science at the *University of Edinburgh*. He graduated from the *University of North Carolina at Chapel Hill* with a major in 'Classics-Civilisation and Philosophy', and a minor in 'History'.

2 For examples see: Reeve (2006), Nussbaum (1979), Ludwig (2002), Neumann (1965).

3 Sedley does give weight to the speech of Agathon, however, he does not attempt to unify Agathon's role with an understanding of *eros* or expand upon the relationship between Agathon and Socrates (2006).

showcase of his rhetorical skill and a list of his own apparent qualities. However, unlike many of the other symposiasts, Agathon plays a consistent role in the dramatic actions throughout the dialogue. Among other examples, he is the host of the banquet, shares a couch with Socrates and Alcibiades, is the object of both their pursuits and is the last one to stay up doing "philosophy" with Socrates. Most importantly, he is the cause and focus of Socrates's questioning and speech at the heart of the dialogue (199c-201e). Based on these thoughts, the focus of this paper is to help establish Agathon as a central character to the deeper philosophical context of the *Symposium*. This interpretation will subsequently help to refocus part of the common understanding of Platonic *eros* from thoughts on individual love to educational methods.

Despite a predominant lack of interest in Agathon, a few studies have utilised him in developing interpretations of the text. In particular, Luc Brisson's 'Agathon, Pausanias, and Diotima in Plato's *Symposium*: *Paiderastia* and *Philosohia*' demonstrates Agathon's role in understanding the dialogue as a pederastic critique. Overall, Brisson differentiates between common pederasty and *Paiderastia*; the latter being a more noble pursuit aimed at *Philosophia*, rather than mere sophistic knowledge or sexual pleasure (Brisson, 2006: 229). He, then, claims that *Paiderastia* is represented in the life long relationship between Agathon and Pausanias. Specifically, Brisson interprets Pausanias's speech as a passionate defense of this noble form of pederasty (Brisson, 2006: 240-45). Further, Diotima's speech develops a critique of *Paiderastia* (Brisson, 2006: 240-51).[4] Overall, Brisson's argument successfully demonstrates the need to incorporate ideas of education, specifically Greek thoughts on the appropriate ways to pursue knowledge, into our understanding of the dialogue. However, despite this success, we will see that Brisson ultimately underplays Agathon's significance to the whole dialogue.

My argument will begin by describing Agathon's role in the drama of the dialogue. Specifically, the first section will examine Agathon as a historical character and what effect that has on our understanding of Plato's literary characterisation. Further, the section will examine how particular parts of his biography have led previous interpretations to dismiss Agathon's importance in the dialogue. Finally, it will argue for a positive and substantial reading of Plato's Agathon that highlights his good-nature and moral intuitions. In the second section, we will turn to a close examination of Agathon's speech. Overall, the speech reflects moral and rational inclinations, but also youth, naivety and a lack of philosophical refinement. The third section will, then, demonstrate that Socrates develops both his questioning and speech as a response to

---

4 In addition, see (Nightengale, 1993) for an alternative understanding of the Platonic critique, namely dealing with the encomiastic discourse.

Agathon. On the one hand, Socrates exhibits a level of academic respect for the young man. However, there also seems to be an implicit criticism of the educational system that failed to develop Agathon's potentials. The fourth section will concentrate on the conclusion of the dialogue and argue that Alcibiades comes to embody Athenian education. Further, the focus will be on how the last few pages present a dramatic competition between Socrates and Alcibiades veiled by sexual contact, but actually representing two different educational paths. These pages of the dialogue will also be used to argue for a tragic reading of Agathon that reflects this educational interpretation. Overall, the paper is aimed at reexamining the role of Agathon and including this new understanding of his character to the many intriguing layers of the *Symposium*.

2. PLATO'S CHARACTERISATION OF AGATHON AND THE DRAMA OF THE *symposium*

We should first examine the elements of Agathon's historical biography in order to develop a better understanding of his characterisation in the dialogue. At the *Symposium's* dramatic date, Agathon is around thirty years old (Nails, 2002: 8). At this stage, Greek elite were expected to be married or, at least, no longer the *eromenos* of a pederastic relationship (Brisson, 2006: 233-4). However, Agathon was well-known for having been the lifelong beloved to Pausanias. In fact, this is how Plato introduces both of them in the *Protagoras* (315e). In the comedy *Thesmophoriazusae*, a play produced five years after the dramatic date of the *Symposium*, Aristophanes criticises Agathon for being effeminate and passive (29, 182-5, 217f, 247-50). In a recent interpretation, Peter H. von Blanckenhagen claims that, "In modern slang, Agathon is a drag queen" (1992: 59). Based on this evidence, it seems that Agathon's relationship was both visible to the public and open to some level of criticism. In addition to his reputation for being the constant beloved, Agathon was a successful tragedian. As Plato highlights, the whole reason for this symposium and Socrates's attendance is to celebrate Agathon's first victory in the Dionysia festival (174a). Also, Aristotle credits him with being one of the first tragedians to have developed his own plot: not based on received mythological figures (*Poet.* 1451b, 1454b, 1456a). In 407 BC, he was invited to live at the court of Archelaus, a Macedonian king, presumably under his patronage.[5] Although of a limited nature, these examples of Agathon's prolific playwriting career suggest a more complex picture

---

5 Pausanias was reported to have accompanied him. In addition, this is mentioned both in the *Symposium* (172c) and in Aristophanes's *Frogs*.

than Aristophanes presents. Therefore, we should examine what aspects of Agathon's historical background are highlighted in Plato's depiction.

Turning to the *Symposium*, Plato seems to develop a character that reflects the multiple historic dimensions of Agathon. Throughout the dialogue, Agathon is consistently referred to as the youth of the group, despite the fact that he is a fully matured man (223A). Furthermore, in the *Protagoras*, Agathon is mentioned for his youth and good looks, as well as, his relationship to Pausanias (315d-e). Therefore, Brisson rightly emphasises that Agathon assumes the role of the beloved among the symposiasts. The following passage clearly highlights how Agathon, despite his age, was still functioning in a pederastic mindset, "Socrates, come lie down next to me. Who knows, if I touch you, I may catch a bit of the wisdom that came to you under my neighbor's porch. It's clear you've seen the light. If you hadn't, you'd still be standing there" (175D). As Brisson argues, this passage "associates the transmission of knowledge from one man to another to that of a seminal fluid" (Brisson, 2006: 229). More specifically, Brisson demonstrates how he is part of the noblest or highest form of male-male relationships, in that his and Pausanias's pursuits are aimed, first and foremost, at *Philosophia*. For Brisson argues that Agathon's main purpose in the dialogue is his connection to Pausanias, and together, they subsequently develop a defense of *Paiderastia*. However, this is also where Brisson's essay demonstrates its limited perceptions of Agathon's character.

In analysing Plato's characterisation of Agathon, we should also notice how Plato emphasises Agathon's intellectual and moral intuitions completely independent of his relationship with Pausanias. Returning to the *Protagoras*, Plato describes Agathon as the young and attractive boy that shares a couch with his lover Pausanias. Yet, in the *Symposium*, from the moment Agathon enters the scene, he demonstrates a certain moral intuition that is not at all connected to Pausanias. For example, when Aristodemus shows up uninvited, Agathon graciously welcomes him and even provides an excuse for why he did not invite him in the first place (174e). As the host, Agathon also demonstrates a serious intellectual interest, as is evident in his choice to invite certain symposiasts. Mainly from the upper echelons of Athens, the group includes members of the aristocracy (Phaedrus and Eryximachus), and the intellectual elite (Aristophanes and Socrates) (Von Blanckenhagen, 1992: 60). Overall, this suggests that Plato clearly sought to portray Agathon in a positive light. Agathon is interested in intellectual pursuits and demonstrates a willingness to develop such attempts.

In addition to this positive characterisation, Plato develops an important relationship between Agathon and Socrates. First, they manage to share a couch with each other, despite the latter's late arrival. If the em-

phasis was on Agathon the beloved, as Brisson argues, why does he not share a couch with Pausanias, as he did in the *Protagoras*? This decision was certainly not an accident. It seems clear that Agathon specifically held the seat next to him for Socrates. In the passage quoted above (175d), Agathon reveals the reasons behind this desire, namely a pursuit of Socrates's wisdom. On the one hand, this statement reflects a pederastic background, by associating knowledge with physical contact. However, the decision also reflects a larger interest in the pursuit of knowledge. Pausanias does not seem to be offended by Agathon's decision, suggesting that Agathon's intentions are noble and intellectually driven. The statement should also be compared to that of Alcibiades's, later in the dialogue (218e-219a). This connection demonstrates how they both share an interest in gaining knowledge from Socrates, as well as the pederastic model they utilised for acquiring such wisdom. In addition, Socrates responds to both in a rather ironic way. On the one hand, Socrates claims his wisdom is but "a shadow in a dream" compared to Agathon's own wisdom (175e). On the other hand, Socrates claims that Alcibiades is trying to exchange a bronze standard of wisdom, "the merest appearance of beauty", for a gold standard, "the thing itself" (218e). However, it seems that Agathon understands Socrates's sarcasm and even responds with some of his own, "Now you've gone too far, Socrates" (175e). This playfulness suggests a certain level of respect between Agathon and Socrates that, as we will see, is emphasised throughout the dialogue.

This relationship between Socrates and Agathon also seems to structure some of the main philosophical and dramatic developments of the dialogue. Firstly, Agathon's speech is the only one to be given a more traditional Socratic refutation. Further, Socrates structures his entire depiction of Diotima's advice in relation to this interaction with Agathon, "I think it will be easiest for me to proceed the way Diotima did and tell you how she questioned me. You see, I had told her almost the same things Agathon told me just now" (201E). While this statement will be examined in detail later for its philosophical importance, on a basic level, it deepens the importance of the connection between these two characters. In the final scene of the dialogue, Plato depicts a comedic display that once again emphasises this connection between Socrates and Agathon, "As if the real point of all this has not been simply to make trouble between Agathon and me!" (222D). Finally, of all the other characters, Agathon stays awake the longest doing "philosophy" with Socrates. Therefore, based on this evidence, it appears that Plato purposefully structured major pieces of the dialogue around their relationship.

Overall, the combination of Agathon's central role in the dialogue and his moral intuitions suggest some preliminary conclusions. Firstly, Agathon attempts to foster an environment focused on attaining wisdom, and he is particularly interested in what Socrates has to say. Further, it seems that by highlighting the relationship between Agathon and Socrates, we can begin to see some of the educational components to the dialogue. Therefore, we should try to comprehend how the connection between them affects the *Symposium's* philosophical program. Also, we should see what this interpretation of Agathon adds to the dialogue's conception of *eros*. Nevertheless, it seems that based simply on this examination of his character and relationships, Agathon has merited a closer analysis.

3. AGATHON'S SPEECH

In this section, we will examine how Agathon's speech helps develop the philosophical progress of the *Symposium*. For the most part, modern interpretation has dismissed his speech as simple and self-absorbed. Brisson describes the speech as "empty but magnificently constructed" (Brisson, 2006: 245). He argues that it is a showcase of his training in Gorgianic rhetoric and ability to use poetic citations, rather than a speech aimed at real intellectual pursuits. This criticism is furthered by Nehamas and Woodruff in their translation of the dialogue. They suggest that the speech really only reflects Agathon's connection to the god Eros, his youth, and physical beauty. For example, after he concludes his speech "everyone there burst into applause, so becoming to himself and to the god did they think the young man's speech" (198a). Here, Nehamas and Woodruff suggest that the 'to himself', "refers to him as the youngest and best-looking man present" (Nehamas et la, 1989: 37). Overall, interpreters seem to disregard Agathon's speech on account of his claim that Eros is happy, young, beautiful and delicate, (the reflection of the beloved). This sort of focus on individual beauty suggests that Agathon's speech is a digression from the intellectual atmosphere of the *Symposium* (Brisson, 2006: 245).[6] Thus, it does not seem to merit any further analysis.

However, despite these valid criticisms, it does seem that Agathon's speech reflects the same moral intuitions and intellectual potential that were established in the previous section. Firstly, we should draw our attention to comments Agathon makes just before he begins his speech: "Why, Socrates, you must think I have nothing but theater audiences on

---

6 This is also present in Von Blanckenhagen (1992: 62).

my mind! So you suppose I don't realise that, if you're intelligent, you find a few sensible men much more frightening than a senseless crowd" (194B). This reflects a common Socratic argument where "wisdom is the only really good thing and ignorance (lack of wisdom) the only really bad one. Because the majority are unwise, they cannot reliably produce the effects they want" (Reeve, 2002: 64). Specifically, the statement echoes the *Crito*, "But, my dear Crito, why do we care so much for what most people think? For the most reasonable men, whose opinion is more worth considering, will think that things were done as they really will be done" (44c-d). Although Agathon's statement does not directly refer to this Socratic idea, it does indicate that he, at the least, has an appreciation and respect for knowledge over mere approval of the masses. This example will help to establish a trend within Agathon's speech, namely that while Agathon has the right initial tendencies, his philosophical background is clearly at a novice level.

At the start of his speech, Agathon continues to develop a very "Socratic" approach to his praise of Eros, "what he is like (referring to Eros), no one has spoken about that" (195A). Further, he says, "Now, only one method is correct for every praise, no matter whose: you must explain what qualities in the subject of your speech enable him to give the benefits for which we praise him" (195A). Here Agathon establishes a philosophical methodology for his speech. On the one hand, this scheme for praising *eros* seems to refer back to some of the earlier Platonic dialogues, where Socrates is pursuing his famous definitions of piety or other virtues and trying to understand the what-it-is (*ti esti*). In fact, Agathon uses very similar language to the earlier Socratic dialogues. Further, Socrates then approves of this method later in the *Symposium*, even using the same *hoios* language multiple times, "Indeed, Agathon, I thought you led the way beautifully into your speech when you said that one should first show the qualities of Love himself, and only then those of his deeds. I must admire that beginning. Come, then, since you have beautifully and magnificently expounded his qualities in other ways [...]" (199c). This indicates that Agathon develops a correct method for understanding *eros* in his speech. However, since his philosophical training is not fully developed, he does not use the *ti estin* terminology.

On the other hand, this approach also seems to differentiate his speech from the speeches before his; Agathon is interested in the what-it-is, whereas the other speeches are interested in the benefits of Love for humans. For example, Aristophanes thinks that love should be praised for drawing people together (192e). In addition, this method seems to separate Agathon from some of Socrates's other interlocutors throughout Plato's works. For example, in *Meno*, Socrates asks Meno what virtue is (71d). However, Meno responds by listing a number of instances of

virtue, "There is virtue for every action and every age, for every task of ours and every one of us" (72a). In this example, we should notice that Socrates is using the more technical *ti esti* language. Nevertheless, it seems Agathon should be acknowledged for correctly establishing the task for Socrates's more in-depth and advanced philosophical discussion.

In addition to proposing a proper philosophical method, Agathon also develops a legitimate explanation of *eros* that focuses on its moral qualities. He claims that Eros is the most beautiful and the best of the gods. However, the god is not just young and beautiful, but also just, courageous, temperate and wise (196d). Overall, he sets up a picture of *eros* in its ideal state. In doing this, the speech has also produced the criticism's that were discussed earlier. Yet, Agathon also proposes that when loving, one is inspired towards the good and becomes happy and peaceful. For example, after Apollo was touched by *eros*, he "invented archery, medicine, and prophecy" (197a). In contrast, those who are not guided by *eros* end in oblivion (197a). This perspective on *eros* does seem to be rather naive and hopeful, but it does not seem to be mere fluff or self-absorbed.

However, in developing this naive perspective on *eros*, Agathon also demonstrates his lack of philosophical understanding. He suggests that when *eros* is involved the person is always directed to do good things, "That too is how the gods' quarrels were settled, once Love came to be among them; love of beauty, obviously, because love is not drawn to ugliness [...] But once this god was born, all goods came to gods and men alike through love of beauty" (197b-c). This picture is somewhat accurate, but only insofar as the love is directed to the good. It seems we could think of *eros* directing one's passions at bad goals. Perhaps a simple example is the case of Alcibiades and how his passions resulted in the virtual destruction of the Athenian empire (Wohl, 2002). The main problem with Agathon's speech is that it fails to appropriately define and defend *eros* in its entirety. Overall, his speech leaves open many unanswered questions that Agathon fails to recognise. Frisbee Sheffield well illustrates some of these questions in *Plato's Symposium: Ethics of Desire*. For example, how and why does *eros* engender the creation of or the possession of good and beautiful things? His answer is that since it is all of those good and beautiful things, its presence in other beings promotes those tendencies (196e). However, this runs into conflict with his notion that it is creative. If one already has these qualities within them, why should they bother with these pursuits at all? (Sheffield, 2006: 25). While there seems to be an easy solution to this question, namely that someone would want to create more of it, Agathon fails to account for this possibility. In addition, when Socrates refutes him on this topic, he

fails to utilise this answer in his defense. Overall, these criticisms seem to highlight both his naive goodness and a highly unrefined philosophical education.

Following these examples, it seems we can draw some preliminary conclusions regarding Agathon's larger role in the dialogue. On the one hand, Brisson suggests that "his speech echoes that of Pausanias, which is intended as a defense and illustration of *paiderastia* as an educative instrument that enables the achievement of excellence in all its forms, particularly in the area of poetry his Eros possesses all the virtues and can transmit them to everyone" (Brisson, 2006: 246). In fact, Agathon's speech does demonstrate an impressive use of meter, poetic citations and understanding of Gorgian style rhetoric.[7] However, Agathon's speech also reflects true philosophical progress on a number of important levels. These intellectual features of his speech and his overall set of moral intuitions suggest that Agathon represents more than just the product of his relationship with Pausanias. Further, in the next section of the dialogue, Socrates picks up on this philosophical progress both in his refutation of Agathon and in Diotima's speech. Therefore, an important question arises: how can we make sense of Agathon's speech within the larger context of the dialogue? It seems that his role is two-fold. On the one hand, Agathon does express a certain goodness that is reflected throughout the dialogue. However, on the other hand, these qualities are corrupted by his education. Throughout the speech, Agathon is consistently disrupted by his need to include references to poetry or utilise his rhetorical training, "I am suddenly struck by a need to say something in poetic meter" (197c). Further, we can see this complex role expressed within Socrates's refutation and Diotima's speech. Socrates demonstrates a level of respect towards Agathon, but there is also a serious criticism of him, as the product of Athenian education. Therefore, within this criticism, we can now see that even the "fluff" of Agathon's speech serves a larger purpose in the dialogue's philosophical development, namely a critique of Agathon's educational upbringing.

4. AGATHON, SOCRATES AND DIOTIMA

In the following section, we will see how Agathon's philosophical role helps develop the content of the more debated sections of the dialogue. To start, despite the large number of speeches presented before Socrates's turn, the philosopher decides to focus his response around and towards

---

7 Roberts provides a full discussion of the influence of Gorgias on Agathon's rhetoric. 1900.

Agathon. As we have seen, Socrates approves of Agathon's method-ological approach (199c). Then, he "corrects" one of the major flaws of Agathon's argument by demonstrating Agathon's inability to defend his definition of *eros* against dialectic (199d-201c). However, unlike many Platonic dialogues, Socrates follows this elenchus by admitting to make the same mistakes Agathon has just made (201e).[8] This can then be read in context with his introduction of Diotima, "I shall try to go through for you the speech about Love I once heard from a woman of Mantinea, Di-otima [...] She is the one who taught me the art of love" (201d). So that Diotima's speech becomes a mouthpiece for how Socrates was able to move out of the same state of *aporia* that Agathon now possesses. Thus, Plato seems to be indicating that Agathon is, at least, on the right philo-sophical track and that we can find partial truths within his speech. Fur-ther, this relationship between Agathon and Socrates helps express the positive and negative educational messages of the *Symposium*, namely an argument for the philosophical or examined life, and the negative critique of current Athenian educational systems.[9] Finally, if Socrates and Diotima are to be Plato's platforms for proposing this argument, then Agathon, the embodiment of both youth and pederasty, as well as intellectual desire and moral intuition, reflects the potential to lead the right type of life and the product of a system that failed to harness that potential.

First, in Socrates's questioning of Agathon, we can start to understand why Agathon is the prime candidate of focus for the instructions. We are not surprised when the refutation is quick and easy; however, it is also the only real "Socratic Dialogue" that we get in the *Symposium*, where Socrates proves the interlocutor to be mistaken by simply showing he believes two contradictory things about the topic. One possible inter-pretation is that the elenchus becomes the first step towards the Socratic concept of the examined life. The next step for the individual is to accept his state of *aporia* and try to resolve it in new ways. At the end of this section, Agathon almost seems perfectly primed to pursue this lifestyle, "I am unable to challenge you. Let it be as you say" (201c). This indicates that Agathon does not become defensive about this refutation, as many of Socrates's interlocutors. In the case of Thrasymachus, in the *Repub-lic*, following his discussion with Socrates, he claims not to be satisfied with Socrates's account and that he could still argue his original point if Socrates would allow him to make a speech about it (350d). In contrast, Agathon claims to be unable to challenge Socrates. This implies that

---

8 Good examples are found in *Euthyphro*, and *Gorgias*.

9 I realise that I have not fully defined these general notions of Athenian education. Nonetheless, I have in mind sophistry, pederasty, etc. Further, it seemed outside the scope of this paper to further develop these notions.

Agathon has no other way of arguing with Socrates. Also, Plato uses the same verb in both Agathon's response to Socrates's questioning, "Let it be as you say" (201c) and Socrates' response to Diotima's questioning, "True, as you say" (202a). While this does not necessarily mean that Socrates's refutation is a good and accurate one, it does suggest that Agathon is a willing participant in the dialectic and brings the argument as far as he can.

Following the elenchus, we should notice Agathon would be in a similar state to Alcibiades when Socrates forced him to "feel shame", (216b) as well as when Socrates was corrected by Diotima for making the same mistakes (201e). Both of these comparisons will help better illustrate the role of Agathon in this central section of the *Symposium*. Firstly, Agathon seems to possess a far superior sense of the importance of wisdom than Alcibiades. Imagine being the "star" of the night and after delivering your speech, the person who you most respect and admire dismantles you in front of your closest friends. Would we expect the same calm and gracious reaction from Alcibiades? Therefore, at least in contrast to Alcibiades, Plato seems to continue to highlight Agathon's moral condition.

Following the questioning of Agathon, Socrates makes a crucial admission, "You see, I had told her almost the same things Agathon told me just now: that Love is a great god and that he belongs to beautiful things" (201e). In terms of the philosophical message, this statement seems to setup Diotima's speech as a response to the problems exhibited in Agathon's speech and even his character. Prior to this statement, Socrates commends Agathon for establishing the right type of methodology, "following your lead, Agathon, one should first describe who Love is and what he is like, and afterwards describe his works" (201d). Both of these statements suggest that Socrates was, at some point, in the same academic position as Agathon is on the night of this infamous symposium. Further, based on this interpretation, Diotima's speech becomes a way for Plato to express how Socrates moved from this level of intellectual potential to the fully actualised embodiment of the philosophical life.

As Brisson has argued, Agathon is involved in the most ideal form of pederasty; a relationship that is defended ardently by Pausanias earlier in the dialogue. But, in addition to this role, it also seems fair to argue that Agathon reflects a wider body of Athenian youth and their education. Perhaps one way of understanding Agathon's role is to think of him embodying the qualities of a youth that might be selected for further education in the Kallipolis of Plato's *Republic*. However, since Agathon is well-beyond the appropriate age to start the training of a philosopher-king, it appears that Plato is emphasising that Agathon's

education failed to fully utilise his good nature and intellectual potentials. Thus, Socrates and Diotima seem to have two tasks set before them: criticise the current educational systems and provide a positive alternative (e.g. the philosophical life).

The speech, however, does not provide the straightforward and direct type of response that we might like. For instance, it is odd that Plato would, in the midst of this "celebrity" event, bring in what appears to be a fictional character. Debra Nails argues that Diotima should be treated on her own as a representation of religion and mysticism and not conflated with philosophy (2006: 193). This sort of reading makes sense with the rest of the characters: a doctor, comedian, tragedian, and philosopher. In addition, we should remember that in the *Symposium* Plato is trying to reach out to a non-philosophical audience. Therefore, he must employ non-philosophical means of getting to a philosophical life. In this sense, Diotima is advising for the philosophical life inadvertently. Perhaps one understanding is that Plato is proclaiming a philosophical or examined approach to the many different topics that are present in the *Symposium*: medicine, tragedy, comedy, religion, etc. Therefore, the Socratic philosopher does not directly need their own speech. Rather, their task is to engage others and force them to participate in dialectical discussions to determine if their theories can lead to true wisdom. This is something that Socrates does directly with Agathon. Therefore, if Agathon is the product of the Athenian educational system, Plato is emphasising to his audience that the system is not fully developing its youth.

In looking at the text, the structure of Diotima's speech seems to follow a very set educative path. She starts by questioning Socrates in a similar way to Agathon's refutation. Then Diotima provides a mythical story for the creation of Eros that defines its role as the product of resource and poverty. Even within this mythical story, Diotima is providing an explanatory argument for her definition of *eros*, something that Agathon was not able to accomplish. In the next section of her speech, Diotima fully fleshes out her understanding of *eros*, "The main point is this: every desire for good things or for happiness is the 'supreme and treacherous love' in everyone" (205d). Firstly, this definition is more in line with Agathon's thoughts of creative passion than with Aristophanes's unique love of another individual. But, she also provides a more comprehensive definition than Agathon. *Eros* should be understood as desire, something that is not tied exclusively to matters of loving other individuals, but also something not tied exclusively to the production of good and beautiful things. Therefore, we should understand *eros* as something that has to be directed in a particular way, if it is going

to assist us in attaining all of the great things Agathon discusses in his speech.

In the next section of her speech, the ascent passages, Diotima develops both a criticism of current forms of education and a positive alternative account that seems to build off of Agathon's speech, refutation, and overall role in the dialogue. Firstly, the critique entails that there is a correct way to use one's erotic passions, and a wrong way. In one particular example, Diotima actually uses the verb *paiderastein* to describe what it means to love rightly, "When someone rises by these stages, through *loving boys correctly*, and begins to see this beauty, he has almost grasped his goal" (211B). This passage suggests that *paiderastein* is a necessary step to "loving rightly", but only insofar as it is used to move up the ladder towards beauty itself. While this example illustrates a critique of pederastic relations, there also appears to be a more comprehensive argument against Athenian education at large. In fact, it seems that *paiderastein* is just an immediate starting point to the ascent. Once one sees beauty within bodies, "The result is that our lover will be forced to gaze at the beauty of activities and laws and to see that all this is akin to itself, with the result that he will think that the beauty of bodies is a thing of no importance" (210c). Therefore, it seems that all sorts of earthly passions fall into stages on the ascent ladder. Thus, we can see that the criticism is that the standard forms of education do not promote the same type of upward movement as the philosophical life. Rather, by simply pursuing one of these educational routes, (rhetoric, politics, poetry, medicine, etc.) there is no incentive to move upward towards the beautiful. Thus, we will need a system that promotes this journey. In the case of Agathon, we see an example of someone who perhaps had the potential to come to the highest mysteries that Diotima discusses, yet, his education did not develop these capacities.

Still, Diotima's speech also develops a positive account of the philosophical life. To move up the ladder and eventually see beauty requires living an examined life, "but the lover is turned to the great sea of beauty, and, gazing upon this, he gives birth too many gloriously beautiful ideas and theories [. . .] until [. . .] he catches sight of such knowledge" (210d-e). In this passage, Diotima draws a distinction between ideas or theories, and fully developed knowledge. Therefore, the ascent to beauty, significantly reflects the examined or philosophical life. Through dialectic, one can turn their ideas into actual knowledge. However, as we have already seen, Agathon is not expected to be able to move up the ladder and see the highest mysteries. Diotima even warns Socrates that he might not move to that level (209e-210a). Nevertheless, it does seem that Agathon embodies the right type of character to which this sort of information would be useful. Even if Agathon is not expected to change

his life after this encounter, the youth that he represents, those with high moral intuitions and good intellectual passions, by reading this dialogue, might learn how to utilise their potential.

5. AGATHON, SOCRATES, ALCIBIADES, AND THE TRAGEDY OF THE *symposium*

In the last section of the *Symposium*, we get to see Plato's drama played out. On one side, you have Socrates (the embodiment of philosophy), on the other, you have Alcibiades (the embodiment of the political and honor-filled life), and, in between, you have Agathon (the embodiment of youth). In this reading, Alcibiades seems to also embody a method for attaining wisdom, e.g. the pederastic model, which cannot be detached from physical acts. Much like Agathon, he sees that Socrates has valuable 'information' that he would like to possess through sexual methods of transference. However, when Alcibiades is forced to "feel shame" by Socrates's refutations, and his weaknesses and shortcomings are revealed, he responds by rejecting philosophy and "caving into (his) desire to please the crowd" (216b). Therefore, we see that Alcibiades's speech serves as a temptation of pederasty and a warning against such Socratic methods of acquiring wisdom. Simply, Alcibiades was once intrigued by what Socrates seemed to possess, yet was, and still is, consistently disappointed. In a way, he seems to represent the common view of Socrates: intrigued, yet, frustrated.

Agathon seems quite content to deny Alcibiades and follow Socrates, "but he won't get away with it; I'm coming right over to lie down next to you" (222E). What is the difference between Agathon and Alcibiades? They are both young, beautiful, intelligent, and successful. Why does it seem that Agathon is eager to accept Socrates's methods? Further, as we have seen, it would not seem plausible that Plato felt Agathon was still in a position to turn towards the philosophical life. Therefore, why would Plato develop this humorous competition over Agathon, especially if his audience knew Agathon did not turn towards the philosophic life?

One of the most humorous sections of the *Symposium* comes in the final dramatic scene between Socrates, Alcibiades and Agathon. On the one hand, Alcibiades is clearly still in love with Socrates, yet, hates him at the same time. Further, the three characters play a game of cat and mouse as they switch seats on the couch. However, as Socrates claims in this last section, "the skillful tragic dramatist should also be a comic poet" (223d). Within this scene, we also find the ultimate tragedy of the *Symposium*. Many people have argued that the *Symposium's* tragic

element exists in the relationship between Socrates and Alcibiades.[10] In particular, we should keep its specific dramatic dates in mind, and view the tragedy as connected to the mutilation of the herms, and profanation of the mysteries that occurred in the following months. As well as, of course, Socrates's death in 399 BC (Nails, 2006: 200). As interpreters have pointed out, many of the *Symposium's* characters were involved in these events that led to the Athenian downfall, namely Alcibiades, Phaedrus and Eryximachus (Nails, 2006: 201). However, this sort of interpretation fails to address the dialogue as a whole. Therefore, by building the dialogue around the relationship between Agathon and Socrates, a new interpretation of this tragedy becomes possible.

After hearing Socrates's speech, demonstrating his own intellectual capabilities and moral intuitions, and resisting the temptation of Alcibiades, Agathon is sitting doing philosophy with Socrates and yet still falls asleep. After this point, it seems fair to assert that Agathon did not drastically change his life. He did not end his relationship with Pausanias, and he continued working as a playwright. Most importantly, he did not pursue a philosophical lifestyle. Certainly, we are meant to place this text in its appropriate historical context, and understand that the majority of the other symposiasts went on to take part in the destruction of their own society. However, in this subtle reading, even the good-natured and morally inclined symposiast is too corrupted by his society to lead the examined life. Following this dialogue, it should no longer surprise us that we begin to see the development of a strict philosophical educational program, specifically with the writing of the *Republic*. Perhaps there is no hope for the self-indulgent Alcibiades, but for all those like Agathon, "the good man", there needs to be a better educational system to guide them towards philosophy. The goal is not necessarily to generate good people out of bad material, but to develop a system that guarantees a full actualisation of one's potentials. The problem with pederasty or rhetoric, even at their most noble forms, are their failures in this regard, namely, truth and goodness are not their main concerns. Therefore, even the good-natured Agathon will not be able to circumvent societal values to achieve an understanding of the good.

BIBLIOGRAPHY

[1] Brisson. 2006. 'Agathon, Pausanias and Diotima in Plato's Symposium: Paiderastia and Philosophia', in *Plato's Symposium: Is-*

---

10 The tragedy of the *Symposium* is further discussed in (Nussbaum, 1979), and (Lear, 1998).

sues and Interpretations and Reception. Lesher, Nails, and Sheffield eds. Center for Hellenic Studies, Trustees for Harvard University. Washington, DC.

[2] Dover, James Kenneth. 1980. *Plato: Symposium*. First Edition. Cambridge University Press: Cambridge.

[3] Grube, George Maximilian Anthony. 2002. *Plato: Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo*. First Edition. Hackett Press: Indianapolis.

[4] Konstan, David., and Elisabeth Young-Breuhl. 1982. *Eryximachus's Speech in the Symposium*. Apeiron 16: 40-6.

[5] Lear, Jonathan. 1998. *Open minded: working out the logic of the soul*. First Edition. Harvard University Press: Cambridge, Mass.

[6] Lesher, James., Debra Nails, and Frisbee Sheffield. 2006. *Plato's Symposium: issues in interpretation and reception*. Center for Hellenic Studies, Trustees for Harvard University: Washington, DC.

[7] Lombardo, Stanley., and Bell Karen. 1992. *Plato: Protagoras*. Hackett Press: Indianapolis.

[8] Ludwig, P.W. 2002. *Eros and Polis*. First Edition. Cambridge University Press: Cambridge.

[9] Nails, Debra. 2002. *The People of Plato*. First Edition. Hackett Press: Indianapolis.

[10] Nails, Debra. 2006. 'Tragedy Off-Stage' in *Plato's Symposium: Issues and Interpretations and Reception*. Lesher, Nails, and Sheffield eds. Center for Hellenic Studies, Trustees for Harvard University: Washington, DC.

[11] Nehamas, Alexander. 1992. 'What did Socrates Teach and to Whom Did He Teach It?' *The Review of Metaphysics* 46, 2, 279-306.

[12] Nehamas, Alexander., and Paul Woodruff. 1989. *Plato: Symposium*. Hackett Press: Indianapolis.

[13] Nehamas, Alexander., and Paul Woodruff. 1995. *Plato: Phaedrus*. Hackett Press: Indianapolis.

[14] Neumann, Harry. 1965. 'Diotima's Concept of Love', in *The American Journal of Philology*. 86: 33-59.

[15] Nightingale, Andrea. 1993. 'The Folly of Praise: Plato's Critique of Encomiastic Discourse in the *Lysis* and the *Symposium*', in *Classical Quarterly*. 43: 112-30.

[16] Nussbaum, Martha. 1979. 'The Speech of Alcibiades: A Reading of Plato's Symposium', *Philosophy and Literature*. 3: 131-72.

[17] Reeve, C.D.C. 2006. 'A Study in Violets: Alcibiades in the Symposium', in *Plato's Symposium: Issues and Interpretations and Reception*. Lesher, Nails, and Sheffield eds. Center for Hellenic Studies, Trustees for Harvard University: Washington, DC.

[18] Reeve, C. D. C., Peter Meineck, and James Doyle. 2002. *The trials of Socrates: six classic texts*. Hackett Press: Indianapolis.

[19] Roberts, W. Rhys.1900. 'Agathon and Aristophanes', in *The Journal of Hellenic Studies*. 20: 44-56.

[20] Sedley, David. 2006. 'The Speech of Agathon in Plato's *Symposium*', in *The Virtuous Life in Greek Ethics*. Reis, Burkhard ed. Cambridge University Press: Cambridge.

[21] Sheffield, Frisbee. 2006. *Plato's Symposium: the Ethics of Desire*. Oxford University Press: Oxford.

[22] Sheffield, Frisbee. 2006. 'The Role of the Earlier Speeches in the *Symposium*: Plato's Endoxic Method?', in *Plato's Symposium: Issues and Interpretations and Reception*. Lesher, Nails, and Sheffield eds. Center for Hellenic Studies, Trustees for Harvard University: Washington, DC.

[23] Sommerstein, A.H. 1994. *Aristophanes: Thesmophoriazusae*. Aris and Phillips: England.

[24] Von Blanckenhagen, Peter. 1992. 'Stage and Actors in Plato's Symposium', in *Greek, Roman and Byzantine Studies*. 33: 51-68.

[25] Wohl, Victoria. 2002. *Love among the ruins: the erotics of democracy in classical Athens*. Princeton University Press: Princeton, N.J.

3 | A NEUROPHILOSOPHICAL AP-PROACH

A Neurophilosophical Approach of the Psychological Process of Chronic Pain

Magali Fernández-Salazar,[1] *Paris IV, Sorbonne University*

1. INTRODUCTION

The only way to tackle the unresolved problem of chronic pain is to study it from different perspectives, that is, a multidisciplinary approach. It seems to me that as philosophers, we cannot talk about pain if we do not take into account other disciplines, (e.g. physiology, neurobiology, cognitive sciences, and psychology) because pain is not only a sensation, nor a perception, but both. The perception of pain by an individual is highly complex and individualised, and it depends on a variety of external and internal influences. The somatosensory cortex is concerned with the appreciation of pain and its quality, location, type and intensity. But, in addition to neural influences, which transmit and modulate sensory input, the perception of pain is affected by psychological and cultural responses.

In this work, I envisage exploring the nature of pain as a typical state of consciousness, while taking account of its physiological and phenomenological aspects, in order to start a discussion regarding the body-mind problem.

The current definition of pain is not complete because it is impossible to identify the nature of pain beyond affirming that it concerns a disagreeable feeling due to a corporal lesion. Some physiologists regard

---

1 Magali Fernández-Salazar is a PhD. philosophy student at *Sorbonne University*. She is also a visiting PhD. student at the *University of Cambridge*. Her current research area is focused in the emotional dimension of phantom limb pain.

pain as a sensation, in which, it is not necessary to take consciousness into account to define the physiological mechanisms. Some philosophers regard it as an intentional mental state without taking into account its neurophysiological aspects to understand its mechanisms in the brain and its relationship with the mind. In addition, some psychologist regard pain from its emotional aspect and somatisation. These approaches are incomplete because of the complex nature of pain, in which, other important aspects are also implicated (e.g. culture). Pain constitutes a legitimate sensation, and a perception too, in the sense where the painful sensation is necessarily integral as a representation. This is to say, consciousness of a bodily lesion is subjected to several levels of comprehension.

Pain, like all other perceptions, can result in illusions, hallucinations (such as a phantom limb), cognitive influences (such as the analgesia of the soldier, or the athlete), and of pathologic elements in which the stimulus is disassociated from representation. The pain of a phantom limb indicates that the brain generates the experience of pain, and that we do not need a lesion to perceive pain, nor a body to feel a body.

With these bases, pain would be the subjective representation of a corporal injury, which includes:

(1) The sensitive element (quale of a painful feeling).

(2) The affective (the aversive emotions which provoke pain).

(3) The volitional (disposing action).

(4) The cognitive (the recognition of the injury, identification and explanation of the perceived feeling).

(5) The behavioral (the movements, the lamentations).

(6) The cultural (the modulation of the painful experience caused by the social and cultural apprenticeship, as well as the influence of personal beliefs).[2]

These elements join together to integrate a complex representation of pain in which each reveals itself as physical and mental.

I shall argue that we cannot consider pain as a determined physiological state (the activation of fiber C conducting the painful information), as it is necessary to explain the quality of the pain, and its phenomenological aspect in respect to the neurobiological mechanisms.[3] Thus, in this document, I propose to study pain as a psychophysical and cognitive mechanism with neurobiological bases, as a subjective, and qualitative experience. In this way, the given neurophysiological

---

2 All of the above are explained in (Díaz, 2007).
3 These neurobiological mechanisms enable us to identify pain in our organism.

and phenomenological should correlate to constitute a psychophysical process. One should be capable to integrate these two perspectives of pain: the objective and the subjective. This approach would consider the psychophysical, and neurophilosophical.

2. MORE THAN THE PERCEPTION OF A SENSATION

Pain is defined by the International Association for the Study of Pain (IASP) as "an emotionally disagreeable sensory experience associated with a damaged tissue past, or potential, or described in terms of such damage" (Pain, 1986: 250). But, even if this definition takes into account the fact that the patient can accept he has no lesion, it seems to me that the definition is insufficient. The "disagreeable feeling" is too vague of an expression to clarify matters. However, this definition is at least consequential, and important, in the sense that it considers pain as something subjective. This is to say, above all, as a state of consciousness, and not just as a sensory modality.

Moreover, this definition considers the experience of pain as implying diverse associations between the sensory and the affective states, which is profoundly aversive. In this sense, pain would not uniquely be a sensation, but also an affective reaction. In other words, pain would not only be a sensation translated into a repulsive emotion, but would, in addition, be a cognitive experience. Nevertheless, this definition, as it stands, destroys the belief some neuroscientists maintain where pain is a noxious sensation exclusively generated by neuronal mechanisms in the brain. The subjectivity of pain does not only depend on internal mechanisms, but also on external sources which could not simply be deduced by a brain without contact with its environment. The emotional dimension participates in the genesis of pain. The culture of the individual, his beliefs, his motivations, and even his economy are implicated in the perception of his pain.

Pain is much more than the perception of a simple sensation. All painful perception has a subjective character that is modulated by the context in which pain intervenes: its meaning, the previous experiences, the culture, even the socio-economic standing of the individual, and his psychological state (anxiety, depression, etc.). But, it does not seem possible to prove a correlation of these elements, nor does it seem possible to bring about an understanding of the process that surrounds the nature of the consciousness of pain. For the time being, we can just try to

bring some ideas, which direct us, gradually, to the scientific discovery of the mechanisms of the consciousness of pain.

The idea would be to seek a theory that could explain the physiological, the cognitive, and the philosophical functions of pain in a unified model.

3. THE PAINFUL CONSCIOUSNESS

Pain is a process in time, in which its different components are linked and mixed in diverse ways, in order to form a whole. This is a complex and distinct representation, which one can call, a painful consciousness. The distinction of the components that would be necessary to make an analysis is not clear because the painful experience is integrated.[4] *A priori*, each of these elements is revealed as physical and mental. As in the case of consciousness, the spatial location of pain can be disconcerting. Irrespective to the type of pain, (acute, chronic, inflammatory, etc.) the subject who normally perceives it, makes reference to the place of the lesion; however, pain is in the brain. This organ fails to be sensitive to injury, which is a strange, and ironic fact.

It is assumed that pain is found in the brain because the painful tracts that leave the nociceptive receptors arrive there, and it appears that it is also in the brain where the feelings, and perceptions, are integrated. Although, we still do not have good evidence, at least not one that is definitive, to indicate how this integration is produced. Furthermore, we still have no evidence to explain in which way the distinctive quality of pain is produced (e.g. its *quale*).

There is always a dichotomy between the phenomenological experience and the scientific evidence: I can perceive a pain in my finger, but in reality, it is in my brain, and not in my finger. In the same way, we could say that one can perceive the outside world, as an exterior, but in reality this world is perceived by our mind. Thus, it is in our brain. "Pain is a representation of a corporal injury in the same way that the visible world is a representation of the space before my eyes, which is constructed in the brain after the transduction by the cones and rods of my retina" (Tye, 1995).

In any case, when one talks about consciousness, whether it be that of pain or the ability to see objects, and all sorts of elements that surround us, it is not easy to make a distinction between objects, mental representations, the *stimuli* which provoke these perceptions, the quality of these experiences, and our social and cultural beliefs. Further, it would seem

---

4 This is due to the fusion of its components.

that it is exactly in the convergence of all these factors that consciousness emerges.

4. THE *qualia* OF THE PAINFUL EXPERIENCE

We should analyse the painful experience. What is the "painful" of pain experience? The answer may be the *qualia*.[5] In other words, the qualitative aspects of the painful experience, the brute and primary sensations of the entire state of consciousness (Hansberg, 2003).

To have an experience is to be in a certain state, or to live through something specific. Can we know how the pain of cancer is if we do not have this disease? We could perhaps imagine it, but we will never feel its effects in order to perceive its pain. This, leads us to affirm that in order to talk about *qualia*, we need to believe in mind and subjectivity. It is difficult to know if a new born could have any perception of pain, or if he just feels a pure sensation, because of its lack of experience. To perceive a pain, we have to have been alive. So, if we consider that the new born had an experience of "life", in the womb, in all the extension of the word, and its implications in the mind and the body, then we could affirm that the baby perceives pain. *Qualia* constitute the most intimate and specific aspect of mental capacity. We do not know its physical basis, and that is why the *quale* is so mysterious and challenging.

While the nervous information has similar characteristics in all the cerebral sectors, (for example the schema of electric discharges is analogous to relatively comparable neurons), the modality of feelings and perceptions is phenomenologically distinct. Common examples of this are the redness of an apple, or the concept of liberty, or tooth pains, or the odor of fresh grass. For the time being, it is impossible for us to comprehend how a physiological phenomenon could have a subjective aspect of something so particular and dramatic as pain, in its various aspects of negative quality, punitive, and noxious consciousness.

But, the most difficult problem to be understood, regarding *qualia* is to know if they are purely a representation. That is to say, if the brute feeling already has a term of reference, or if there is something in *qualia* which escapes functional representation. Further, even if for the time being it is difficult to make a proposition which could be proved with psychobiological experiments, we could possibly consider a psychobiology of *qualia*, but not merely a biological one.

---

5 The plural of *quale*
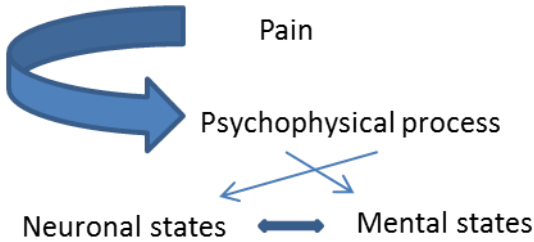
5. THE CORRELATION BETWEEN NEURAL AND MENTAL STATES

One could propose that consciousness could be an emerging property of cerebral activity, and in this case, one could say that pain could correspond to a neuro-mental state. However, such emergence, which would be a functional property, should be correlated to a physiological process that must be well defined. For the time being, we are incapable of such definition. This leads us directly to the problem of mind-body. We must begin to understand the functioning of this emergent process.

In any case, I propose that pain is not only a nervous process, but that it could have a correlation with this process. This is not causality, but a match; reciprocity. This would be:

I avoid the concept of identity and propose one of correlation (with the signs). These two processes could form an independent study, but the result of such studies (of nervous processes) would have to have a significant correlation with the mental states of the individual perceiving pain. This correlation is necessary, as the two aspects, the physiological and the phenomenological, should correspond. There would be neither hard pain, nor a crucial neurological process, without a painful experience.

This formula of correlation opens a door for us to the study of the mechanisms in the consciousness of pain. It could serve to study, not only the aspects of information about pain, but also the sense of the painful quality that should be correlated with the nervous and cognitive aspects of the experience of pain. This would take into account the relation between the qualitative aspect of the mental states of the experience of pain and its neurophysiological aspects.

To arrive at such a stage, it might be necessary to ask which are the specific neurotransmitters involved in the 'painful' process. In other

words, we would need to identify which are the brain areas implicated in this process, the specific structures, their mechanisms, and interactions. One should also find significant correlations with qualitative aspects of the 'painful' experience because we are concerned with the qualitative aspects of the nervous process.

This suggested proposal would be a first step towards the comprehension of the complex processes in the brain. We are not yet able to find a solution to the central problem of the consciousness of pain, but at least, we can find a way to develop a theory of pain by taking into account all the disciplines involved.[6] In any case, it would appear to be undisputable that the psychophysical correlations are preliminary in explaining the qualitative aspects of consciousness, while taking into account the *qualia* of pain as one of the phenomenological aspects in relation to those neurobiological aspects which comply.

BIBLIOGRAPHY

[1] Aydede, Murat. 2001. "Naturalism, introspection and direct realism about pain", in *Consciousness emotion*. 2:29-73.

[2] Craig, A. D. 2003. "Pain mechanisms: labeled lines versus convergence in central processing", in *Annual review Neurosciences*. 26:1-30.

[3] Dennett, Daniel. 1990. "Quining Qualia", in *Mind and Cognition*. Lycan, W. ed. Oxford Blackwells: Oxford. p. 519-548.

[4] Dennett, Daniel. 2008. *De beaux rêves. Obstacles philosophiques á une science de la conscience*. First Edition. de l'éclat Press: Paris.

[5] Díaz, Jose Luis. 2008. *La conciencia viviente*. FCE, Mexico.

[6] Drestske, Fred. 1995. *Naturalizing the Mind*. MIT Press: Cambridge, MA.

[7] Fodor. Jerry. 1990. *A Theory of Content and Other Essays*. Bradford Books/MIT Press: Cambridge, MA.

[8] Folz E.,L., White, L.E. 1962. "Pain relief by frontal cingulotomie". *Journal of Neurosurgery*. 19, 89-100.

[9] Hansberg, Olbeth. 2003. *La naturaleza de la experiencia*. Instituto de Investigaciones Filosóficas. UNAM: Mexico.

[10] Lycan, William. 1996. *Consciousness and Experience*. MIT Press: Cambridge, MA.

---

6 As opposed to work with them separately, in the same area.

[11] Nagel Thomas. 1974. *What Is It Like to Be a Bat?* Philosophical Review 83: p. 435- 450.

[12] Putnam, Hilary. 1997. "The Nature of Mental States", in *Art, Mind and Religion*, University of Pittsburgh Press: Pittsburgh.

[13] Ramachandran, Vilayanur. 1998. *Fantasmas en el cerebro* First Edition. Debate Press: Barcelona.

[14] Ramachandran, Vilayanur., Diane Rogers-Ramachandran, and Stewart, M. 1992. "Perceptual correlates of massive cortical reorganization", in *Science Journal*, 258 (5085): pp. 1159-60.

[15] Reisberg, Daniel. 1997. *Cognition, exploring the science of the mind*. W.W., Northon and Co: NY, USA.

[16] Sabato, Alessandro. 2010. *Dolore basi molecolari*, CIC. Edizioni Internazionali, Roma.

[17] Smart, J.J.C. 1959. "Sensations and Brain Processes", in *Philosophical Review*. 68, reprinted in John Heil: *Philosophy of Mind: a Guide and Anthology*. Oxford University Press: Oxford. p. 116-127, 1959.

[18] Torebjork, H. E., and LaMotte, R., and Robinson, C. J. 1984. "Peripheral neural correlates of magnitude of cutaneous pain and hyperalgesia: simultaneous recordings in humans of sensory judgments of pain and evoked responses in nociceptors with C-fibers", in *The American Physiological Society*.

[19] Tye, Michael. 1995. "A representationalist theory of pains and their phenomenal character", in *Philosophical Perspectives, 9*. MIT Press: Cambridge, MA.

[20] Woolf, Clifford. 1994. "The pathophysiology of chronic pain–increased sensitivity to low threshold A-beta-fibre inputs", in *Current opinion in neurobiology, Elsevier*.

4 | HUME'S SCEPTICISM VS. SWIN-BURNE'S INFERENCE

How Hume's Scepticism Regarding the Design Inference is still Applicable Today

Peter Casurella,[1] *McMaster University*

1. INTRODUCTION

I recently logged into a social networking site to find that a friend had posted a humorous picture for my consideration. Before me, a young man was gleefully engaged in an act of nasal exploration with his index finger. The caption read: "Your finger fits perfectly into your nostril. Checkmate, atheists". This satirical picture is referring to the Argument from Design, one of the classic approaches to arguing for God's existence via an analogous comparison of intentional and ordered objects created by man, to the apparent intentionality and order of the universe; which, it is argued, must likewise have an intelligent creator. In the example of our friend above, the assumption is that it is no accident that one's finger conforms to the size of one's nostril, but rather, that the convenience of this arrangement was thought out beforehand by a god who made man in such a way that objects could be removed with ease from one's nose.

In his *Enquiry Concerning the Human Understanding*, the Scottish philosopher David Hume began what many consider to be the most devastating philosophical attack on the argument from design.[2] His remarks here, combined with his post-humously published attack on the design argument in *Dialogues Concerning Natural Religion*, are considered by many

---

1 Peter Casurella is a second year Masters student in philosophy at *McMaster University*, in Hamilton, Canada. His primary area of research is in the analytic philosophy of religion; specifically focusing on arguments and digressions of interest surrounding proofs for the existence of god.

2 Hume published this paper in 1757.

to have greatly reduced its philosophical importance.[3] However, despite Hume's attacks, the argument has persisted and evolved over the centuries and arguably remains the single most compelling argument for god's existence. In his 2004 edition of *The Existence of God*, the esteemed Oxford philosopher and theologian Richard Swinburne advances a probabilistic form of the Argument from Design, building a case that the design inference leads us to believe that the existence of god is more likely than the alternative. His argument connects with human intuitions in a compelling way. However, he does not succeed in his project. This paper will argue that Hume's 250 year old scepticism regarding the design inference cannot be bypassed by Swinburne's probabilistic attempt at dodging its conclusions. I will begin by reviewing how Hume's epistemological commitments lead him to conclude the irrationality of the design argument. Next, we will examine how Swinburne goes about trying to dodge the problems presented by Hume. Finally, if Hume's epistemology, or something like it, is correct, then I will show that Swinburne's dodge fails to carry his conclusion clear of Hume's scepticism.

2. HUME'S SCEPTICAL PHILOSOPHY AND THE ARGUMENT FROM DESIGN

The design argument is certainly not new. In the 4th century BC, Plato wrote that one of the main reasons for men's belief in the gods was "the order of the motion of the stars, and of all things under the dominion of the mind which ordered the universe" (Plato, 12.966e). Aristotle too believed that the beauty of nature was what first caused men to wonder about how the universe had come to be. He argues in *On Philosophy* that anyone who observes the scope and beauty of nature would "judge both that there exist gods and that all these marvelous works are the handiwork of the gods" (Aristotle). In the middle ages, Thomas Aquinas propelled the argument to lasting fame through his inclusion of it in his five ways of knowing that God exists. Aquinas argues that any non-conscious object, which exhibits purpose in its design, must be under the direction of a conscious and intelligent being. In the same way that arrows do not head towards a bulls-eye without intelligent direction, neither do acorns grow into oaks without intelligent direction. Therefore, everything in nature must be directed towards its goal by someone with intelligence; e.g. God (Aquinas).

---

3 This paper was published twenty-two years later, in 1779.

The Argument from Design really hit its stride in 1804, when William Paley published his book, Natural Theology. In Paley's defense of the rationality of belief, he says that if a man were to find a watch while walking along a beach, and were to observe its intricate parts and ordered operations, even if he had never before seen a watch, that man would surely infer that some intelligence had designed it. Furthermore, if the ordered contrivance of the watch leads us to infer that it must be the product of an intelligent design, then observation of the natural world should lead us to a similar conclusion regarding its origins. This is because, says Paley, "every indication of contrivance, every manifestation of design, which existed in the watch, exists in the works of nature", and what's more, "that the contrivances of nature surpass the contrivances of art, in the complexity, subtlety, and curiosity of the mechanism; and still more, if possible, do they go beyond them in number and variety" (Paley, 1804: 13). If this is true, then we cannot help but be drawn from our observations of a world, which exhibits all the marks of design, to the inference that there must be a designer.

3. HUME'S SCEPTICISM

Hume's attack on this classic argument follows from his epistemological stance on how humans acquire knowledge, and what the limits of our knowledge are. Writing in the same philosophical tradition as John Locke and René Descartes, Hume ascribed what is often called the 'Idea Theory' of human cognition. In brief, Hume's version of the Idea Theory holds that "all the materials of thinking are derived either from our outward or inward sentiment", meaning that every thought which we are capable of, is produced either by impressions from external stimuli (via our five senses) or through the recombination of those impressions in our minds (1778, EHU: 2.5).[4] Ultimately, the entirety of our ideas are derived, directly or indirectly, from sense data which we receive through our physical senses. Therefore, all we can *know* is that which we are able to acquire either from sense data, or derive through a composition of sensory impressions in the mind.

Of central importance is how we are able to form inferences regarding cause and effect. Hume says that we cannot know the fundamental principles behind causation, because this realm of inquiry is entirely closed off to us (1778, EHU: 4.12). Having never before seen a billiard ball, Hume asks the reader to consider how, by simply observing the ball, a person could discover what its effects on other objects might be. It is

---

4 EHU will refer to *Enquiry Concerning Human Understanding*.

only through the experience of watching how the billiard ball interacts with other balls that we are able to discover what its effects on other balls are (1778, EHU: 4.9). No reasonable inferences can be made from the observation of one object towards what its effects might be. Our limited human minds are simply incapable of delivering this kind of knowledge to us *a priori*. Hume, thus, says that it is "in vain" that we should "pretend to determine any single event, or infer any cause or effect, without the assistance of observation and experience" (1778, EHU: 4.11).

The foundation of all our inferences is experience, and these inferences are not founded on any rational processes or reasoning (1778, EHU: 4.14). If our inferences regarding cause and effect were *rational*, then we could observe the billiard ball and deductively reason our way forward to what its effects must be without needing to see it in action (1778, EHU: 4.15). However, since this is not the case, our inferences must be considered *irrational*. All our conclusions regarding 'cause and effect' are based on an expectation that similar causes will follow from similar effects (1778, EHU: 4.20), and this expectation is only the result of the brain becoming accustomed, through constant exposure to the conjunction of two objects or events. In other words, to expect the one to follow from the other, or to expect similar events to follow from similar causes (1778, EHU: 5.4-5.5).

We shall see that it is this expectation of like events, following from like causes, which gives rise to the design inference. Humans have become deeply accustomed to the constant conjunction of machines, which exhibit order, regularity, and purposefulness being created by an intelligent human mind. Therefore, when we see purpose, regularity, and order in nature, our minds are drawn by habit to a similar conclusion regarding the universe as a whole. The fallacy is understandable, but a fallacy, nonetheless.

4. APPLICATION TO THE ARGUMENT FROM DESIGN

After presenting the argument above, Hume begins to follow its consequences to their logical conclusions. In chapter 11 of the *Enquiry*, he advances his doubts as to whether, in light of his version of the Idea Theory, the design inference can be rationally made at all. Hume presents the design argument as an argument from analogy. Complex objects like houses and watches are created by designers. The universe itself seems

to be a complex object analogous to a house or a watch. Therefore, the universe itself must have a designer.

In his later work, *Dialogues Concerning Natural Religion*, Hume allows the assumption that God really is the designer of the universe and goes on to level five arguments against the soundness of the design inference, based on the logical fallout which such an inference leads to. Firstly, he says that the analogy is weak. It is not at all clear that the universe so closely resembles a man-made machine that it always appears ordered or designed. Many inexplicable and seemingly random events occur around us all the time. Next, Hume argues that the argument is also underdetermined. We have no reason to believe that order and complexity only arise from intelligence. "For all we know", says Hume, "matter may have a source of order within it, just as mind does, having it inherently, basically, not acquired from somewhere" (1779, DCNR: 2).[5] Further, Hume argues that if we posit an intelligent designer, that designer must also possess the properties of order and complexity, because causes must contain sufficient properties to produce their effects, and must be alike in enough ways such that the correlation is rational. If God is comprised of order and complexity, then we are simply pushing the query back one stage, and are then, left to wonder what caused an ordered god?

Hume's final two critiques in the *Dialogues* deal with the properties of God which men have ascribed to him. It seems contradictory that a perfect god should create a manifestly imperfect world. It also seems impossible that we could reasonably infer from the universe's creation, any other quality about god, beyond those qualities absolutely necessary to instantiate the universe.

However, Hume also shows that the design inference itself should never be allowed off the ground in the first place. In *Hume, Newton, and the Design Argument*, Robert Hurlbutt nicely summarises this argument:

> The design argument is *not* scientific in that it offers no evidence for the causes of the world order. The world is one particular, not a member of a species, whose members have been observed. In order to demonstrate a cause for any effect it is necessary to have observed the cause and effect in conjunction, indeed, in constant conjunction. And no one has seen the origin of one world, let alone "worlds" (Hurlbutt, 1965: 151).

---

5 DCNR will refer to *Dialogues Concerning Natural Religion*. Hume's comments on this point are almost prophetic. The later development of the theory of evolution, both biological and stellar, would later reinforce this point of his very strongly. In many ways, matter *does* seem to organise itself, at least sufficiently, and for long enough, to produce beings with the kind of complexity which we exhibit.

On Hume's understanding of cognition, causal inferences only arise from cases where we have observed a constant conjunction between objects, such that the mind becomes accustomed to expect one object to follow upon observation of the first. Every time we see a house being built, we have also observed people actively building such things. The conjunction of the two objects is very firmly fixed in our minds: intelligent purpose and designed object. This is the case with all the objects of human contrivance. Since our minds come to expect this conjunction, whenever we see what appear to be objects, or systems, which seem to exhibit order, complexity, and intent, we are naturally drawn to infer that a designer must be the cause.

However, while it may be psychologically understandable why we make the inference, this does not justify our application of this inference to the universe as a whole. It is only "when two *species* of objects are found to be constantly conjoined, that we can infer the one from the other" (1778, EHU: 11.30). In the case of the instantiation of the universe, we have only observed one instance of an effect without even observing its cause. Therefore, we cannot make any rational conclusions about what the cause of the universe must be.

5. SWINBURNE'S BAYESIAN APPROACH

In his influential book *Inference to the Best Explanation*, Peter Lipton describes how these inferences work. When we find ourselves wanting to explain a certain phenomenon, but insufficient data exists to deductively describe the event, "given our data and our background beliefs, we infer what would, if true, provide the best of the competing explanations we can generate of those data", so long as the best is good enough for us to make any inference at all (1999: 58). Given competing explanations for some event *a*, we compare and contrast rival explanations and determine which one is more likely to be true based on its ability to account for the phenomena.[6]

Since Hume's 18th century attack on the design argument, and in light of the purchase his thought has gained in the intellectual community, many of the more respected attempts to salvage the design inference have adapted by making IBE claims.[7] In 2004, the distinguished Oxford philosopher and theologian Dr. Richard Swinburne published an updated edition of his book, *The Existence of God*. In this book Swinburne argues that although there are no good deductive arguments for

---

6 We also see if it has good fit with other observed phenomena and general rules.
7 Hereafter, let 'IBE' stand for 'Inference to the Best Explanation'.

the existence of God, a successful "P-inductive" argument can be built to show that the existence of God is more likely than the alternative, and therefore, is the best possible explanation for the universe (2004: 12-13).[8]

Acknowledging that the traditional starting points of the Argument from Design are problematic, in light of both Hume's scepticism and the development of Evolutionary Theory, he instead sets out to simply argue that it is more *probable* that a designer god exists than not, due to the evident existence of 'spatial order' and 'temporal order'. By spatial order, Swinburne is referring to the apparent arrangement of nature in orderly and purposeful ways, such as the structure of the human eye. The character Cleanthes in Hume's *Dialogues* sums up the idea of spatial order nicely.

> Consider, anatomise the eye; survey its structure and con-
> trivance, and tell me, from your own feeling, if the idea of
> a contriver does not immediately flow in upon you with a
> force like that of sensation (1779: DCNR.3).

Not only does Swinburne see spatial order in the arrangements of objects, but he considers nature to be like a 'machine-making machine,' in that it is constructed so as to give rise to the order which we see in the eye. He supposes that the theory of evolution can be fully incorporated into the design inference as a part of the intricate and ordered operations of the universe, and argues that it is highly improbable that humans would have evolved without an intelligent creator, god, to set up the universe-machine to produce them.

A key supporting point for this argument is Swinburne's supposition that the universe has been 'fine-tuned' to allow for the development of life. In brief, there are several fundamental forces in the universe which govern the interaction of matter, and if the relative strengths of those forces had been different by the smallest degree from what they are now, then cosmic evolution would have followed a different path, and human life would never have been able to evolve (Swinburne, 2004: 172-190). In light of the great number of alternative ways the universe could have been, Swinburne proposes that it is more likely than not that there is an

---

8 A 'P-inductive' argument is one in which the premises add to the probability of the conclusion, and a *correct* P-inductive argument is one in which the premises make the conclusion more probable than its negation (Swinburne, 2004: 6). Swinburne examines the probability of the conclusions of several arguments, for and against, the existence of god. He measures their relative strength in terms of the extent they confirm the hypothesis that God exists. Swinbure intends to show that if h is the hypothesis that God exists, and eb is the evidence from a particular argument for God's existence, then P(h b1+b2+b3 [...] bn) > P(h b1+b2+b3 [...] bn). Therefore, it is more likely than not that God does exist.

explanation that *this* universe exists, and not another. The most likely explanation is a theistic god.

Swinburne thinks that his second argument, from temporal order, is the strongest of the two. By temporal order, he is referring to "regularities of succession", or "patterns of behavior of objects", such as the perceived passage of time, the laws of physics, and the 'physical laws', by which we are able to predict that certain future events will follow from certain causes (2004: 151). Swinburne argues that even more than spatial order, the idea that the world is characterised by a "vast pervasive order" lies at the core of the design inference (2004: 155). It seems very unlikely that we could exist at all without the universe being characterised by these pervasive regularities. Cognition seems to require it, as does our day to day survival. Even the fundamental laws which give rise to evolutionary development seem to require this kind of temporal order (2004: 158).

Swinburne rejects two objections to his claim that the apparent temporal order is significant. The first objection is that it is the human mind which imposes order on the universe in order to meet the human need for prediction and control. The second objection is that it should not seem remarkable to us that the universe evinces order, because we would not be here to question it if the right kind of order did not exist in the first place (2004: 156).

In response to the first objection, Swinburne argues that it is the order of the universe itself which is a necessary condition for our minds to have evolved to their present state. If this is true, then even if human minds do impose some order on the universe, there must first have been certain fundamental regularities already operating to have allowed minds, which recognise order, to arise at all (2004: 156).

In response to the second objection, Swinburne appeals by analogy to our ability to recognise when any given event is improbable. If you were to flip a coin fifty times and see a head come up on every toss, the barest knowledge of probability would inform you that you had just observed an extremely improbable event. Now, if someone were to say to you, "I will flip this coin fifty times and the moment a tail is flipped, I will shoot you dead". Upon seeing the fiftieth heads come up, you should not be surprised because you could not have seen anything else and still be conscious of it. Swinburne says that it is absurd to maintain that we should *not* still think the situation highly improbable. Following the analogy, it is therefore, perfectly rational to be surprised that we exist in a universe like ours, given how many fortuitous roles of the cosmic were needed for us to be here at all.

Swinburne's move here is an interesting twist on an old line of reasoning known as the Anthropic Principle. The Anthropic Principle states

that we should not be surprised to find ourselves existing, because if things had been much different in the universe, we would not be here at all to reflect upon the improbability of our existence. This line of reasoning is often deployed against the theistic assertion that our surprise at finding ourselves here should lead us to infer that someone purposely designed the universe to allow for our existence. Swinburne argues that his alternate interpretation should be the preferred one because there appears to be far more order in the universe than would merely be required for the existence of humans (2004: 156).

In summary, the phenomena of temporal order makes the existence of god more probable because, of all the ways the universe could have turned out, it happens that there do exist laws and regularities that explain how we could have developed, and how we are able to then "extrapolate from past to future events with normal success" (2004: 164). It is more probable that a proper explanation exists to account for this than not.[9]

In essence, what Swinburne proposes is that there are two possibilities: (1.) There is order and regularity in the world because there is a god with all the sufficient properties to make it so. (2.) There is order and regularity in the world and this is simply where the explanation stops. Swinburne concludes that from both, the properties of spatial and temporal order, (which we observe) that it is more probable for (1) to be true; as opposed to the universe being simply, a brute fact.

6. BAYES'S THEOREM MISAPPLIED

While well constructed and intuitively appealing, Swinburne's argument is flawed. In this final section I will demonstrate how his probabilistic side-step fails to take the design inference clear of Hume's sceptical attack, by showing how the key analogy as to how we might draw inferences regarding god, cannot be reasonably accepted.

To begin with, Swinburne's overarching case for the existence of god is a Bayesian approach, meaning that it makes use of Bayes's Theorem to calculate and compare conditional probabilities; which serve, to render a judgment as to which condition is more likely to have produced the observed effect. Bayes's Theorem is used to render as true, any hypothesis which a certain body of evidence confirms as probable, or more likely than alternative hypotheses (Joyce). So, if we have a hypothesis (h), then given certain evidence (e), the probability of (h) given (e) is greater than (h) given (e). This is commonly expressed in the following way:

---

9 The proper explanation being god.

P(h e) > P(h e)

In order to judge whether or not the design inference adds or detracts from his overarching case, Swinburne needs to decide if the probability of god's existence, given the evidence from design, is greater or less than 0.5. This figure is then incorporated into the overall evidence for his cumulative case. As we have seen, Swinburne concludes that the probability of god existing, given the evidence we have that spatial and temporal order exist, is greater than the probability that temporal and spatial order could exist without god existing. Is he justified in reaching this conclusion?

Swinburne's approach may seem very subjective, but this is precisely what the Bayesian approach is meant to address. The approach allows one to make intuitive and general claims about the relative probabilities of various causes. Thus, we can make a claim about which is likely to be the actual 'explanation', out of a pool of competing explanations. So long as one's conclusions are reasonable enough for most people to accept them, then general probabilities can be assigned to cases. Subsequently, these general probabilities can be added together to build a strong evidential case for whatever hypothesis is the most likely candidate.

For example, take Swinburne's discussion of the fine-tuning of the universe. Naturally, there is no way to calculate how many possible calibrations of the fundamental forces of the universe would have allowed cosmic evolution to give rise to life. However, it does seem reasonable to assume that there are a great many more ways in which the cosmic forces could have been arranged, such that human life never could have arisen. If the number of situations which would have *not* given rise to human life is judged to be greater than the number of situations which *would* have, then we should conclude that the possibility of human life *not* arising is statistically more probable. Given that we find ourselves in the less likely state of being alive (in an apparently ordered universe), then we are justified in thinking that it is probable that some explanatory hypothesis (which explains our current state) is true; in contrast to just assume that we have simply beat the odds.

Think back to the example of the coin toss. If we find ourselves alive after the 50th toss, we might reasonably presume that the coin itself was not fair. Perhaps both sides of the coin were imprinted with a head. It seems more probable that there is some further explanation, which might account for the unlikely series of coin tosses, other than dumb luck. In thinking of the unlikely scenario in which we find ourselves in the universe, the god hypothesis explains why we find ourselves here in much the same way as the unfair coin hypothesis explains why the man with the gun did not shoot me dead. The point is simply that, if one was

asked to make an educated guess based on the data available, one will either bet in favor, or against, the existence of god. For Swinburne, the unlikelihood of finding ourselves here means that we ought to allot the balance of probability in favor of the existence of god. If the probability is only a fraction more convincing in favor of the god hypothesis than not, then, we are justified in reasonably inferring that god really does exist.

7. THE PROBLEM WITH SWINBURNE'S APPLICATION

No scientist would object to occasional uses of Bayesian calculations, in fact such calculations are part of our day to day life, and can even be descriptive of the foundations of very successful science. Mackie admits that "we are justified in arguing inductively, in extrapolating observed regularities to unobserved cases" (1990: 147). A beautiful example of such successful reasoning was the discovery of the planet Neptune by German astronomer Johann Gottfried Galle. Galle discovered Neptune after several other astronomers had independantly predicted its existence based on irregularities in the orbit of Uranus. Galle theorised that these irregularities were best accounted for by the gravitational influence of a postulated, yet undiscovered, eighth planet. The prediction that Neptune existed before it was discovered can be couched in Bayesian terms as:

> The probability that Neptune exists, given the evidence of Uranus's orbit and our understanding of gravitation, is greater than the probability that Neptune does not exist.

There is a key difference between this example of a very successful IBE and Swinburne's argument for god's existence. In the case of Neptune, there was a probable cause, which fell within the realm of normal science, which could potentially be investigated. If I am allowed to employ a bit of my own probabilistic reasoning here, it seems that it is far more probable that an unobserved cause which operates *within* our universe can be discovered, than that we could discover a cause *of* the universe itself.

I might propose, for example, that the cause of the current universe was a trans-dimensional, super-alien sneezing, and that the sneezing of such aliens always produces explosions of space-time that have ordered regularities as one of their inherent features. Now, it seems unlikely that we would ever be able to investigate if my sneezing-alien theory is cor-

rect, or rather, if the god hypothesis is correct. Since the conditions of these possible causes are beyond the physical universe, then, the truth of these matters lies beyond our ability to investigate them as well. However, this is simply an *ad ignorantiam*, and there is more we can say on this issue.

When building an IBE argument, what makes an explanation 'best', "is always relative to the available competitors it faces, meaning that IBEs always involve comparative evaluations of evidential support among competing hypotheses" (Ratzsh). Swinburne argues that what makes god the best explanation is that the god hypothesis provides the simplest explanation for the observed phenomenon (the universe); while, it provides the greatest explanatory power when compared to other hypotheses (2004: 82). We are justified in believing that this hypothesis is probably correct if "any gain of explanatory power would be outweighed by a corresponding loss of prior probability", and if "any gain in prior probability would be outweighed by a corresponding loss of explanatory power" (Swinburne, 2004: 82). The god hypotheses, argues Swinburne, is the simplest hypotheses available, because all of the conditions necessary for the instantiation of the universe are available in one explanatory step; the actions of God. Positing any further beings beyond God gains us no explanatory power. Nonetheless, why this hypothesis has greater simplicity and explanatory power than competing explanations such as the sneezing super-alien theory, is very puzzling.

In both theories, we have a practical end to investigation. We do not possess the proper investigatory tools to look beyond the proposed entity, in either case. It is not possible to discover any information about whether or not there is an explanation of god or the super-sneezing alien, and thus, our investigations must end with the postulation of some such being as an explanation for the universe. The sneeze of the alien has all the conditions necessary to create our universe, just like the proposed properties of God. Both theories postulate only one being with sufficient properties to create the current universe, and any other properties of such a being are things which we can only speculate at.

Swinburne might argue that the infinite attributes of God, his omnipotence and omniscience for example, would count as more simple properties than a super-alien with some definable set of finite properties. Nevertheless, I suggest that the burden of proof rests on Swinburne's shoulders to show how some infinite set of properties that are sufficient to create an ordered universe are any more simplistic than some undefined set of finite properties, which are equally sufficient for creating an ordered universe. If both are fully sufficient for instantiating the universe, how can we, from our limited perspective within the universe, pass judgment on which scenario is more likely? The properties of the

alien's sneeze are sufficient for creating the universe, but we can not reasonably infer anything else about these properties except their sufficiency. By the exact same reasoning, although we might speculate that God's properties are infinite, all we can reasonably infer is that they are equally sufficient for the instantiation of the universe. Beyond this limit, we are at a loss, and the possible simpler infinity of god's characteristics will not help us adjudicate which theory is correct.

It seems that we cannot rationally adjudicate which theory has either, greater explanatory power, or prior probability based on our current evidence. The only advantage that the design hypothesis has over the sneezing alien theory is that it connects well by analogy with our intuitions regarding purposeful design. But if something like Hume's point regarding our inability to make inferences where we have no experience is correct, then we have no rational basis for judging which of these two causes is more likely. We cannot know, then, if the analogy holds at all. Thus, the analogy should not influence our adjudication of the issue. For it to count in his favor, Swinburne would need to first show us that knowledge of the domain beyond the physical universe can somehow be accessed.

As a final point, I wish to point to the evidence of stellar and biological evolution, which Swinburne himself does not reject, as a counterexample to Swinburne's desire to extend the intelligent design analogy to account for the universe as a whole. At this point in our scientific understanding of the world, we have a very good picture of how selective processes can lead to the refinement of matter into intricate systems. From the evolution of the solar system down to the adaptations of white and black moths in Great Britain to avoid predation, we see many examples of natural processes mimicking intelligent processes. It is not clear at all that the apparent order of nature really is order. John Mackie argues that although we associate the products of human invention with human intelligence, "we have no good *empirical* reason for taking the 'marks of design' as marks of design" (Mackie, 1990: 144). In fact, we have no empirical evidence at all, since the realm of which we wish to make causal claims of, is forever beyond our ability to observe. The intuitive drive for why we should consider the design theory a better explanation than its competitors is, therefore, not intuitive at all, but actually a presumptuous mistake.

8. BRINGING IT BACK TO HUME

In Hume's *Dialogues*, the character Philo, (speaking for Hume himself) argues: "Our ideas reach no further than our experience. We have no

experience of divine attributes and operations" (1779, DCNR.2). Therefore, Hume asserts that we can have no ideas regarding divine attributes and operations, including ideas about the origin or design of the universe. For Swinburne's argument to side-step Hume's scepticism, he would have to show, not only that the god-hypothesis is the most probable cause of a universe that appears to be ordered, but also that we are justified in:

(1) Identify what the causes of the universe might be.

(2) Assign relative probability values to these causes.

(3) Export our inferences from the physical realm (in which such processes of reasoning are useful), to a realm which we are not justified in believing. And, see if this has any resemblance or commensurability with the rules and regularities of the observed universe.

Swinburne entirely fails to do any of these and his Bayesian instincts are, thus, unjustified.

9. CONCLUSION

If belief in the design hypothesis is so irrational, why is it that so many people choose to believe it? Hume offers his error theory through the lips of Philo, saying that people choose to accept the design argument because it fits in nicely with their already existing web of beliefs. Moreover, they wish to continue holding those beliefs. Given the harshness of life, the idea that there exists a perfect, eternal, and good creator who will give humankind eternal happiness in an eternal 'hereafter' makes the trials of this current life more bearable. Hume admits that the arguments in favor of the design hypothesis are psychologically compelling, but they are not in themselves, numerous or forcible (Hurlbutt, 2012: 165). Hume prescribes modesty to sooth the nerves of those who must abandon the design hypothesis in light of his scepticism, saying that by discovering our own limitations in these sorts of investigations, "we may make a kind of merit of our very ignorance" (1778, EHU: 4.14). Swinburne's probabilistic attempt at doing an end-run around Hume's scepticism simply does not dodge far enough.

BIBLIOGRAPHY

[1] Anonymous. 2012. Available on the World Wide Web at <http://i2.kym-cdn.com/photos/images/newsfeed/000/282/360/62b.jpg> [Accessed: 12/03/10].

[2] Aquinas, Thomas. 1265-74. *Summa Theologica*. Edited by Timothy McDermott. 1998. Christian Classics Press. Allen: Texas. USA.

[3] Aristotle. 350 BC. *Physics.* Translated by Robin Waterfield. 2008. Oxford University Press: Oxford.

[4] Hume, David. 1778. *An Enquiry Concerning Human Understanding*. Oxford Philosophical Texts. 1999. Oxford University Press: Oxford.

[5] Hume, David. 1779. *Dialogues Concerning Natural Religion*. Penguin Classics. 1990. Penguin Group Press: London.

[6] Hurlbutt, Robert. 2012. *Hume, Newton, and The Design Argument*. University of Nebraska Press: Lincoln.

[7] Joyce, James. 2008. 'Bayes' Theorem', in *The Stanford Encyclopedia of Philosophy* Available on the World Wide Web, at: http://plato.stanford.edu/ar theorem [Accessed: 27/03/12].

[8] Lipton, Peter. 1999. *Inference to the Best Explanation*. Second Edition. Routledge Press: London.

[9] Mackie, John Leslie. 1990. *The Miracle of Theism*. Sixth Edition. Clarendon Press: New York.

[10] Paley, W. (839. *Natural Theology*. Hathi Trust Digital Library. Available in the World Wide Web at: <http://catalog.hathitrust.org/Record/0057762

[11] Plato. UD. *The Laws*. Translated by Trevor Saunders. Penguin classics. 2004. Penguin Group Press: London.

[12] Ratzsh, Del. 2012. 'Teleological Arguments for God's Existence', in *The Stanford Encyclopedia of Philosophy*. Available on the World Wide Web at: <http://plato.stanford.edu/archives/spr2012/entries/teleologica arguments> [Accessed: 28/03/12].

[13] Swinburne, Richard. 2004. *The Existence of God*. Second Edition. Oxford University Press: Oxford.

5 | RIGHTNESS, BLAMEWORTHINESS, AND THE DOCTRINE OF DOUBLE EFFECT

Caroline Lyster,[1] *McGill University*

According to consequentialist moral theories, the right thing to do is whatever brings about the best consequences. While this does seem to be the way that we make decisions in the real world, a strict application of consequentialist reasoning can sometimes lead to our allowing, or even doing, harm to others in order to bring about good. This seems intuitively wrong. We need action guiding principles to prevent this from happening; principles that will require us to help people and, at the same time, prohibit us from harming them. But, a strict application of principles like beneficence, or non-malfeasance, can also lead to counterintuitive results in cases where the seemingly 'right thing to do' results in collateral damage. What is going on here? Either, we must further modify our action guiding principles, or our intuitions about these collateral damage cases are wrong. I believe it is the latter; in these cases we are confusing intuitions about what is right with intuitions about what is not blameworthy. This means that the doctrine of double effect, the principle that is invoked to explain the apparent permissibility of collateral damage, is not a deontic one, but an aretaic one.[2]

In order to understand where the doctrine of double effect comes from, we should first understand what it is invoked in response to. Con-

---

2 The deontic principle points toward what is right, and the aretaic tells us whether or not a person should be blamed for what they have done.

sequential moral theories are concerned with outcomes, and the right thing to do is whatever leads to the best consequences. On these theories actions have only instrumental value; that is, they are not good or bad in themselves, they are only good or bad insofar as they bring about good or bad consequences. The answers given by consequentialism are usually in line with our intuitions about cases where moral decisions must be made. Consider the following example:

> Transplant 1: A doctor has access to five organs for transplant and has two choices for how to distribute them. He can (1) transplant all of them into one patient who is suffering from the failure of multiple organ systems or (2) transplant each organ into one of five people who are all suffering from failure of a single organ.

According to consequentialism, the right action would be the second option, where the organs are used to save the maximum number of people, and therefore, do the maximum amount of good. Most people would intuitively agree that this is the right choice.

But, if it is all about the consequences, what is the right action in this second transplant case?

> Transplant 2: A doctor has five patients, each dying from the failure of a single organ. One day, a healthy young person comes into the clinic for a simple check-up. As it turns out, this healthy young person is a match for all five of the doctor's dying patients.

If actions are right or wrong only insofar as they bring about good or bad consequences, then the right action in this second transplant case should be to kill the healthy patient and distribute their organs to the dying five. However, this seems intuitively wrong.

Our intuitions about the second transplant case demonstrate the need for forward-looking, or deontic, principles that can judge certain actions as right or wrong. Since most consequentialists would accept some version of the principle of beneficence, this seems the logical place to start, so action guiding principle number one should be that we should act in ways that help others. Beneficence would explain why we feel the way we do about the first transplant case: the right action is to distribute the five organs to five different people because this action helps more people. A second deontic principle can be deduced from our intuitions about Transplant 2: obviously the reason it is wrong to kill the healthy patient in order to save five others is because to do so would be to harm them. Therefore, action guiding principle number two may be something like non-malfeasance: do not only act to help others, but also act in such a way as to not harm them.

Non-malfeasance as a guiding principle is certainly a good thing, but we should further define what exactly is ruled out when we say, "do no harm". The most straightforward way of interpreting this statement would be as a prohibition against actively doing harm as opposed to merely allowing it to happen, and this distinction does capture why we feel the way we do about the two transplants: in the first transplant case, the one person is merely allowed to die whereas in the second, the healthy patient must be killed. However, as James Rachels has shown elsewhere (1975), the distinction between doing and allowing harm is tenuous, and ultimately seems to amount to not much more than a desire to keep our own proverbial hands clean. So, however we formulate it, our principle of non-malfeasance must account for harms that arise from intentional omissions. In other words, we need a deontic principle that accounts for the intuitions we have about situations like this one:

> The Trolley and the Fat Man: You and your tour group are sightseeing in a trolley yard when you spy five people tied to the tracks some way down from the platform you are standing on. Then, you notice that, an out of control trolley is rolling towards them. Among your tour group members is a fat man, and out of the corner of your eye you spy the tour operator sneaking up behind him with the obvious intention of pushing him in front of the trolley to save the five on the tracks.

If the principle in question simply prevented us from harming others, we could let someone else push the fat man in front of the trolley, but our intuitions here seem to indicate otherwise. A better formulation of non-malfeasance would be something like what Warren Quinn proposes in his version of the doctrine of doing and allowing (1989). In here, when we say, "do not harm", we prohibit both positive and negative agency, and actually mean something more like, "do not harm or deliberately fail to prevent harm". This version of the principle prohibits sitting back and letting someone else push the fat man in front of the trolley, while still permitting the non-deliberate harm that occurs in the original transplant case.

These straightforward deontic principles of beneficence and non-malfeasance appear to capture our intuitions about why certain actions are right or wrong, but there are instances where a strict application of them would have counter-intuitive results. These ones, are cases where it seems as though doing the right thing will result in collateral damage. For example:

> Classic Trolley: You are walking along, minding your own business, when you come across five people tied to the trol-

> ley tracks. Then, you notice that, an out of control trolley is careening down towards these five people and will kill them if you do nothing. You spy a switch, and discover that if you were to flip it, the trolley would be diverted onto a sidetrack, saving the five. However, there is a single person tied to this sidetrack, and if you divert the trolley this person will die (Foot, 1978).

What should you do in this case? Most people would say that they should flip the switch, but that action would violate the principle of non-malfeasance, which requires that we act in such a way as to not harm others.

The fact that we have this intuition indicates that our current deontic principle of non-malfeasance is insufficient. We need a way to prevent harm from coming to the healthy patient in Transplant 2, while at the same time allowing for diversion in Classic Trolley. The solution that has often been proposed for this problem is the doctrine of double effect (DDE), which would further refine our basic principle of non-malfeasance. According to this version of the principle, when we say, "do no harm", what we actually mean is: "do not harm or deliberately fail to prevent harm in order to benefit someone else". In other words, the kind of malfeasance that is prohibited by the DDE is that where someone is benefitted at the expense of someone else. This modification to non-malfeasance allows us to take an agent's intentions into consideration; when we use the DDE to evaluate an action, we ask ourselves whether the agent performing it is aiming toward the harm being cause, or whether that harm is merely a side effect of the action that will be done. Using this principle would allow diversion in the Classic Trolley case because the five people are not benefitted at the expense of the one. Rather, the single person's death is a mere side effect of their salvation. The DDE also gets us the results we want in the transplant cases. In the first case, we are allowed to give the organs to five people, and not just one, because that person's dying is not our intention. In Transplant 2 the doctor is not permitted to kill the healthy patient because he would intend that person's death.

While the DDE seems plausible at the out set, it is not without problems. Foremost among them is this: why should an agent's intention be the slightest bit relevant in determining the right thing to do? Saying that intention matters would mean that the right action, would only really be so, if the agent did it for the right reasons. But, if that agent has the wrong intention, what might have been the right choice, would suddenly become the wrong one. To illustrate this point, consider a slightly different version of the classic trolley case:

> Paul and the Trolley: You are walking along, minding your own business, when you come across five people tied to the trolley tracks. Then, you notice that, an out of control trolley is careening down towards these five people and will kill them if you do nothing. You spy a switch, and discover that if you were to flip it, the trolley would be diverted onto a sidetrack, saving the five. However, there is a single person tied to this sidetrack, and if you divert the trolley, this person will die. Upon closer inspection you discover that the single person is your mortal enemy Paul. You have no real desire to save the five, but because you really hate Paul, you decide to flip the switch.

In the Classic Trolley our intuitions indicate flipping the switch is the right thing to do, but as soon as we flip the switch in the new scenario it feels as though we have done something wrong. In some cases, not only can considering intention make the right action wrong, it can also make the wrong action right. For example:

> Mom and the Trolley: You are walking along, minding your own business, when you come across five people tied to the trolley tracks. Then, you notice that, an out of control trolley is careening down towards these five people and will kill them if you do nothing. You spy a switch, and discover that if you were to flip it, the trolley would be diverted onto a sidetrack, saving the five. However, there is a single person tied to this sidetrack, and if you divert the trolley, this person will die. Upon closer inspection, you discover that the single person is your mother. You have nothing against the five, but decide not to flip the switch in order to ensure your mother's survival.

If the DDE is right, and what you should do in a given situation is linked directly to your intentions, then there are three different right actions in each of these trolley cases: (1) it is right to divert as long as you want to save the five, (2) it is wrong to divert if you only want to hurt the one, and (3) right to not intervene if the one is a person whom you want to save. But, can this actually be the way things work? Is it really only right to divert in some cases? Such a notion is inconsistent with the very idea of objective moral rules, and a good deontic principle should be true, regardless of the person performing the action, or the particular situation in which they find themselves.

A second problem with the DDE is this: when is one 'close enough' to intend harm so that it ought to be prohibited by the principle of non-malfeasance? We would like our deontic principle to account for all of

our intuitions about collateral damage, and there are certain intuitively wrong situations that seem to be permitted by the DDE. For example:

> Poison Gas Hospital: You are a doctor working in a hospital. Five of your patients are dying, and the only way to save their lives is to manufacture a certain chemical that will cure their disease. Unfortunately, the gas that is created as a by-product of this synthesis is poisonous, and the lab in which the chemical will be synthesised vents into a neighbouring room. This room is home to a single person who, for some unknown reason, canot be moved (Foot, 1978).

If we acted in accordance with the DDE, we would be allowed to manufacture the chemical to save the five because the death of the single person in the attached room is an unintended side effect. However, our intuitions seem to suggest that this would not be the right choice; there is something about manufacturing the chemical and releasing a poison gas into a neighbouring room that seems awfully close to intending harm. Even though we do not intend to harm the single person, in manufacturing the chemical we set into motion a series of events, which we know will culminate in their death. Perhaps, then, we could say that a person performs an action that is "close enough to intending harm" when they initiate a series of events with full knowledge of the harmful consequences that will result. Given this, our deontic principle of non-malfeasance should be further refined to say something more like this: when we say, "do no harm", we really mean "do not act, or deliberately fail to act, if your doing so will (a) cause someone undue harm, or (b) cause harm to one person in order to benefit another, or (c) set into motion a series of events through which you know someone will be harmed". This version of the principle explains why we may not manufacture chemical to save the lives of the five dying patients in the Poison Gas Hospital case. It also explains why we may not stand back and watch while someone else pushes a fat man in front of the trolley, and why we may not kill one healthy person in order to save five others.

You may have noticed, we turned to the DDE to try and make sense of a single intuition: that of being permitted to divert in Classic Trolley. But, in order to make the DDE work we needed to define exactly what it is prohibiting. Further, we could determine that setting into motion a series of events, which culminate in a foreseen harm, is close enough to intend harm for warrant being prohibited. But, this modification has a strange consequence: it does not allow diversion in the Classic Trolley case. The act of diverting the trolley to save the five sets into motion a series of events resulting in the foreseen death of the single person on the sidetrack. Therefore, this act is prohibited by the principle of non-malfeasance, as it has been articulated.

Let us take stock of the moves that have been made to get us to this point. Firstly, there was consequentialism which, while attractive in it's quest for the most desirable outcome, can lead to problems where agents may be allowed to cause a great deal of harm to one person in order to create good for more. This is intuitively wrong, so there must be a principle of non-malfeasance to balance out the ethical requirement to do good things. In attempting to spell out exactly that this principle of non-malfeasance might amount to, we came to the conclusion that saying, "do no harm" means something more like "do not directly cause or deliberately fail to prevent harm". The problem arose when we realised that this version of our principle of non-malfeasance does not allow for the right action to include collateral damage. We attempted to account for this intuition by invoking the DDE, but quickly found that any useful version of this principle is eventually not going to allow for diversion.

Perhaps, instead of changing our principle of non-malfeasance to account for the Classic Trolley intuition, we should turn to the case and figure out exactly what it is that we are responding to. Logically speaking, if what we are reacting to is the permissibility of diversion, then any case of diversion will yield the same intuitive result. To test this, consider this case, first presented by Peter Unger (1996):

> Trolley on a Hill: You are walking along, minding your own business, when you come across five people tied to the trolley tracks. Then, you notice that, an out of control trolley is careening down towards these five people and will kill them if you do nothing. You spy a switch, and discover that if you were to flip it, the trolley would be diverted onto a sidetrack, saving the five. However, the sidetrack abruptly ends at the top of a hill, and if you divert the trolley, it will end up rolling down the hill and killing an innocent person reading a book in their backyard.

The only difference between this case and that of the Classic Trolley is the location of the innocent bystander: in the latter case, the five are tied to the trolley tracks: however, the single person is sitting in the backyard, some distance away, minding his own business. But, this difference seems to be enough to change our intuitions about diverting. In the Classic case, we want to flip the switch, but in the new case, the answer is not as clear. Is the distance, the diverted trolley must travel to harm the single person, really a relevant factor in determining whether flipping the switch is the right thing to do? It seems arbitrary to say that it is permissible to flip the switch in cases where the bystander will be killed soon, and not in cases where the bystander will be killed later.

Diversion fails the test, so there must be something wrong with our intuitions about the Classic case. We might be able to take a cue from

Peter Unger, who suggests that a mental phenomenon called "projective grouping" can affect the way we think about different cases, and subsequently, how we react to them. According to Unger, "often we view a certain serious problem as being a *problem* for only those folks viewed as bring (grouped together) in a particular situation" (1996: 97). So, the location of the single person in each case makes a difference as to whether or not we group them with the five. In the case of the Trolley on the Hill, the single person sitting in the backyard is not close enough to be counted in the group of people that have a problem. Conversely, in the Classic case, the single person tied to the sidetrack is seen to be as threatened by the runaway trolley as the five. In other words, when we picture the Classic Trolley case, we think of it as looking something like this:

> Two Runaway Trolleys: You are walking along, minding your own business, when you come across two sets of parallel trolley tracks. Five people are tied down to one set, and there is a single person tied down to the other. Then, you notice that, two out of control trolleys are careening down both sets of tracks. You spy a bazooka with only one round, and determine that you can safely blow up one of the trolleys without harming anyone.

This is a case where all six people are clearly in imminent danger, and we may, in this case, be permitted to use our resources to save five rather than one. However, in the Classic case, the single person on the sidetrack is not in any danger until the switch is flipped. We create the danger to him, and given this, our principle of non-malfeasance will not allow us to divert the trolley. But, even if we are careful to keep in mind that the single person, either on the sidetrack, or in the yard, are not in imminent danger, there is still a part of us that wants to flip the switch. Why? I propose that this intuition has nothing to do with what is right, and everything to do with what is blameworthy. That is, we may feel that diversion is okay because we would not blame someone for making that decision.

Whilst deontic principles look forward to tell us the right thing to do, aretaic assessments look back on choices that have been made, and determine whether or not the agent who made the choice is blameworthy. Now, blameworthiness is only somewhat related to whether or not a particular action was right or wrong; when we assign blame, there are other factors that we may take into consideration. One aspect of assigning blameworthiness is the relative wrongness of an action; that is, an agent may due more or less blame depending on exactly what they have done. Consider the example of stealing. It is always wrong, but a person who steals a car may be more blameworthy than someone who steals a

pack of gum, and this relative wrongness plays out in the way that the penalties for stealing are assigned by the justice system. A person stealing a car would be more harshly punished than a person stealing a pack of gum. This difference in 'sentencing' is directly related to how much blame we feel that each person is due.

A second component is responsibility. That is, the more an agent is responsible for their action, the more they are blameworthy. For instance, an agent who kills someone by jumping on them from some height can be blamed more than someone who is pushed and falls. We are also tempted to view people as less blameworthy if they act, and appeared to be right, given the information that they had. Additionally, we do not blame people for failing to do the right thing, when it would be very difficult for them to do so. Unfortunately, none of these aspects can explain our blameworthiness intuitions in the Classic Trolley case, as diverting the trolley, and killing the tied-person, are definitely on the bad side of relative wrongness.[3] But, we still want to save the five, so there must be another reason why the choice is not blameworthy.

I believe that this is a good place for the doctrine of double effect to come back into play. As a deontic principle, the DDE had problems, but as an aretaic assessment, it may be able to help explain why we want to divert the trolley; though it is the wrong thing to do. If we recall, the DDE states something like: all else held equal, there are stronger reasons against bringing about harm as a means to an end than there are against bringing about harm as a side effect. Used to assign blameworthiness, it would allow us to say that an agent would be due more blame for harms they intend to cause than they would be for harms that they bring about by accident.

Using the DDE in an aretaic way can help us explain why we find it so attractive, even though it is apparently flawed. One of the most attractive features of the DDE is that it takes an agent's intentions into account, and though it was demonstrated that reasons for performing a certain action should not have an effect on whether that action is right or wrong, there is reason to think that an agent's intention has something to do with his blameworthiness. Consider a variation of a case we have seen before:

> Transplant 3: A doctor has five patients, each dying from the failure of a single organ. One day, a healthy young person comes into the clinic for a simple check-up. As it turns out, this healthy young person is a match for all five of the doctor's dying patients. But, since one of the people dying

---

3 We also tend to think that the person who flips the switch has no one to blame but himself.

from organ failure is the doctor's mortal enemy, he decides
to let the single healthy person walk away unharmed.

In this case, the doctor does what we would intuitively believe to be
the right thing, but he does it for the wrong reasons. Even though his
action is right, we are averse to the choice, indicating that he is blame-
worthy in the death of the five.

So an agent's intention is probably a third component of blamewor-
thiness, but how does this relate to the DDE, and the diversion cases?
It cannot be as simple as saying that someone is less blameworthy for
diverting the trolley because they intend to save the five. If this were
the case, then that same person would not be blameworthy for pushing
the fat man off the platform, so long as they intended to save the people
tied to the tracks. So the DDE's usefulness as an aretaic principle can-
not be directly tied to the result that an agent wants to achieve through
his actions, or they would be able to justify cases like the Fat Man, or
Transplant 2, on the basis that their intention is to save the five. The
aretaic DDE must be able to separate cases of intentional harm from
those of collateral damage, or else we run the risk of falling back into
consequentialism.

The question is: what makes cases of collateral damage different? Or,
what is it that makes our intuitions about blameworthiness in the Classic
case different from those in a case like the one Judith Jarvis Thomson has
proposed (1985):

> The Loop: You are walking along the trolley tracks, mind-
> ing your own business, when you come across a fork. You
> notice that some way down, the tracks come back together,
> and just after that, there are five people tied down. Then,
> you notice one out of control trolley. If you do nothing, the
> trolley will turn left at the fork before rejoining the other
> track and killing the five. However, if you flip the switch
> the trolley will instead take the right hand track. There is a
> fat man tied to this track, and if the trolley hits him, it will
> not continue on, to kill the five.

The difference between the cases is not the intention of the agent who
does (or does not) flip the switch; in both cases the goal is to save five
people. The difference between the cases is the role that the damage
plays. That is, whether or not the death of the single person is necessary
for the agent to achieve their goal. In the Loop case, the five can only
be saved if the trolley is diverted into the fat man; if he was not there,
flipping the switch would have no effect on the fate of the five. The
trolley would just loop back onto the main track, and kill them. In
contrast, the death of the single person in the Classic case is purely

coincidental; their presence on the sidetrack does not necessarily figure into the agent's calculations. Perhaps our aretaic form of the DDE is better off without any strict appeal to intention, and ought to instead be worded like this: it is more blameworthy to require that someone be harmed in order for your goals to be achieved than it is to cause someone to be harmed as a side effect of your actions.

Of course, the DDE is always going to be subject to the problem of closeness, and the aretaic form of it is no exception. The problem of closeness brings to light the enquiry of when, would something actually be prohibited, or in this case made blameworthy, by the DDE. The principle, (as I have given it above) enquires what exactly harming someone in order to achieve your goals looks like. Earlier in this paper, the Poison Gas Hospital was given as an example of a case that is intuitively wrong, but would nonetheless, be permitted by the DDE. Opponents could bring up that case again, and point out that the DDE still permits it, though in this case that permission only has to do with the manufacture of the chemical and the consequent death of the single person not being worthy of blame.

My response is this: the critics are right, and as an aretaic principle, the DDE will say that the doctor's choice to save the five in the Poison Gas Hospital case is not blameworthy. But, even though we may not blame the doctor for his choice, that lack of blame does not make the action right. In contrast, if the manufacture of the chemical in question required the death of the one; for example, if it could only be made within that one's body, this would be an example of a blameworthy (and wrong action). Closeness is not such an issue for the aretaic version of the DDE, because this assessment is not trying to talk about right or wrong. The aretaic DDE is only meant to help explain our intuitions about cases involving collateral damage, and why sometimes we may want to do things even though they are wrong.

We arrive to the conclusion. We began with consequentialism, and saw that concentrating only on achieving the best outcome can lead to choices that seem intuitively wrong. We need forward-looking deontic principles to help us make choices. The first is easy to accept: the principle of beneficence requires that we act in ways that help others. But, we cannot only concentrate on doing good things; we must also try to avoid doing things that are wrong. So, we must balance beneficence with a principle of non-malfeasance that prohibits us from either causing, or deliberately failing, to prevent harm. Nonetheless, these principles fail to account for the intuitions that we have about cases of collateral damage like the Classic trolley, where we want to divert harm (even though someone else will be harmed by our doing so). This is where the doctrine of double effect was first introduced, but it ultimately failed as an

action guiding principle. It does, however, work as an aretaic one; our intuitions about collateral damage can be explained in terms of diverting not being blameworthy. Ultimately, the death of the ones on the sidetrack is not a necessary part of our plan. It is important to note that the action of diverting still violates the principles of non-malfeasance, it is still wrong; but, other factors mean that we would not blame a person for choosing to flip the switch.

BIBLIOGRAPHY

[1] Allowing, *The Philosophical Review* 98 no. 3. 287-312.

[2] Foot, Philippa. 1978. *Virtues and Vices*. Basil Blackwell: Oxford.

[3] Jarvis Thomson, Judith. 1985. 'The Trolley Problem', in *The Yale Law Journal* 94 no. 6. 1395-1415.

[4] Quinn, Warren. 1989. 'Actions, Intentions and Consequences: The Doctrine of Doing and Allowing', in *The Philosophical Review* 98 no. 3, 287-312.

[5] Rachels, J. 1975. 'Active and Passive Euthanasia', in *New England Journal of Medicine*. 292. 78-80.

[6] Unger, Peter. 1996. *Living High and Letting Die: Our Illusion of Innocence*. Oxford University Press: New York.

University of
St Andrews

600
YEARS