

8-13-2019

Kant and Joseph Butler on Autonomy, Moral Obligation, and Stoic Virtue

Samuel Munroe

Follow this and additional works at: https://scholarworks.gsu.edu/philosophy_theses

Recommended Citation

Munroe, Samuel, "Kant and Joseph Butler on Autonomy, Moral Obligation, and Stoic Virtue." Thesis, Georgia State University, 2019.
https://scholarworks.gsu.edu/philosophy_theses/261

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

KANT AND JOSEPH BUTLER ON AUTONOMY, MORAL OBLIGATION AND STOIC
VIRTUE

by

SAMUEL MUNROE

Under the Direction of Eric Wilson, PhD

ABSTRACT

Scholars have compared Joseph Butler and Immanuel Kant's moral theories, claiming that they both center on the concept of autonomy. In this thesis, I argue that, despite this superficial similarity, they disagree about the core commitments of their conceptions of autonomy. Butlerian autonomy relies on inferring from the normative authority of conscience to the descriptive that human nature is adapted to virtue, and from this descriptive claim about human nature to moral obligation. Kant rejects these inferences, and therefore rejects the key components of Butlerian autonomy. Moreover, I argue that Kant's rejection of Butler's conception of human nature renders Butler's stoic conception of virtue unavailable to Kant.

INDEX WORDS: Kant, Joseph Butler, Autonomy, Moral Obligation, Stoic Virtue, Moral Philosophy, History of Philosophy, History of Ethics

KANT AND JOSEPH BUTLER ON AUTONOMY, MORAL OBLIGATION AND STOIC
VIRTUE

by

SAMUEL MUNROE

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2019

Copyright by
Samuel Patrick Munroe
2019

KANT AND JOSEPH BUTLER ON AUTONOMY, MORAL OBLIGATION AND STOIC
VIRTUE

by

SAMUEL MUNROE

Committee Chair: Eric Wilson

Committee: Timothy O'Keefe

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2019

DEDICATION

For Sophia, who knows the demands of conscience

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Eric Wilson, whose support, encouragement, and insight has improved both the quality of this thesis and the quality of my life throughout my time at GSU. I am also indebted to Tim O’Keefe, whose comments on the penultimate draft of this thesis were immensely helpful. The intellectual and social atmosphere at Georgia State has been supportive and stimulating, for which I thank my fellow graduate students. I am especially grateful to Botian Liu for many helpful and enjoyable conversations about Kant and Butler and for his thoughtful comments on various drafts this thesis. I also want to thank everyone who participated in our thesis workshop, namely, Nick Alonso, Hunter Cantrell, Andrew Gilley, Chase Halsne, Spencer Knafelc, Nick Linhares, and Emily Tilton. And I owe a special debt of gratitude to Matt Kelley and Lucas Opgenorth, both of whom have been excellent friends, roommates, and conversational partners. The friendship of Marisa Maccaro has also been proven invaluable throughout the final stages of writing this thesis. For sparking and fostering my interest in philosophy, I am eternally grateful to my philosophical community at MSU Denver, Sarah Adair, Nick Barnes, Alexa Brown, Haley Burke, Per Hovem, Paul Jacobs, Joey May, Corey Polster, Luke Straka, and especially James Reid. Finally, I want to thank my parents, Tom Munroe and Katherine Steets, and my sister, Emily Munroe, for their love and support in all aspects of my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	V
1 INTRODUCTION	1
2 KANT ON EMPIRICAL MORAL PHILOSOPHY	3
3 BUTLER ON OBLIGATION AND HUMAN NATURE	9
4 KANT ON REASON AND HUMAN NATURE	17
5 KANT AND BUTLER ON HUMAN NATURE	21
6 BUTLER ON STOIC VIRTUE.....	25
7 KANT ON STOIC VIRTUE.....	30
8 CONCLUSION	37
REFERENCES.....	40

1 INTRODUCTION

When Immanuel Kant and Joseph Butler are mentioned in the same breath, it is usually for the sake of claiming that they are, in some general sense, engaged in a highly similar project. Both are interested in justifying claims about moral obligation, and both argue that human beings are obligated to act in one way rather than another, at least in part, because we provide ourselves with a law that commands us to do so. For Butler, this means that starting with “a matter of fact” about “the particular nature of man” one can demonstrate the obligation to follow conscience and act virtuously (P: 12).¹ Kant argues that the absolute authority of reason’s moral law grounds obligation. For both Kant and Butler, then, grounding moral obligation need not (or, in Kant’s case, cannot) involve referring to mind-independent facts about things such as God’s will or a harmonious order of nature. Instead they aim to show that moral obligation can be grounded in what we might loosely refer to as the structure of practical agency. It has thus been claimed by scholars such as Darwall that Kant and Butler are engaged in a common project insofar as they both provide an ethical theory centered around what we can call, using Kant’s language, *autonomy* (Darwall 1993: 232- 41).²

Moreover, Butler’s account of moral obligation and autonomy is tightly bound up with his account of virtue, which scholars agree is inspired by the ancient Stoics.³ In the preface to his *Fifteen Sermons*, Butler states that he aims to explain and defend the “inward feeling” of the “ancient moralists” that they expressed in the claim that “man is born to virtue, that it consists in following nature, and that vice is more contrary to this nature than tortures or death” (P: 13). It

¹ References to the *Fifteen Sermons* are to sermon and paragraph number, according to *Fifteen Sermons and Other Writings on Ethics*. “P” refers to the preface to the *Fifteen Sermons*. References to Butler’s *A Dissertation of Virtue* are to paragraph number, preceded by “D”.

² Butler himself never uses the term ‘autonomy,’ opting instead most frequently for the claim that as man “is a law to himself,” but it fairly characterizes his view (P: 29; 2: 8; 3: 5).

³ For an especially helpful analysis of Butler’s commonalities with ancient Stoicism, see (Long 2003).

has also been argued that Kantian moral theory bears a strong semblance to ancient Stoic moral theory. In particular, his insistence that only those maxims that can be integrated as a law into a harmonious natural order are morally permissible has rightly been seen as similar to the Stoic imperative to harmonize one's soul with nature.⁴ One might, then, fairly think that Kant and Butler's accounts of virtue will bear some strong similarities to one another as a result of their common Stoic influence.

But how similar are Kant and Butler's moral philosophy, really? To what degree can one assimilate Butlerian autonomy to Kantian autonomy, or *vice versa*? And are the reasons for claiming that Kant and Butler are both strongly influenced by Stoic morality sufficiently similar for establishing some commonalities between Kant and Butler? Autonomy occupies a central position in the history of modern ethics, and Kant and Butler are two of its earliest and most prominent proponents. Moreover, Butler's account of virtue follows directly from his arguments for autonomy. Getting clear about the relationship between their respective conceptions of autonomy, and its role in grounding obligation, then serves a two-fold purpose. The first is to better understand the core commitments of Kant and Butler's moral theories, and especially their accounts of virtue. Second, and at least as importantly, it should elucidate the parameters within which autonomy became a central category in moral philosophy. My hope is that bringing Kant and Butler into conversation with one another can, indirectly at least, cast new light on our contemporary (Kantian) concept of autonomy by providing us with one that, I contend, stands in stark contrast with it, namely, that of Butler.

In this thesis, I argue that, despite some surface-level agreements between Kant and Butler, they in fact disagree about many of the core commitments of their moral theories.

⁴ (Reich: 1939)

Although Kant and Butler are both interested in autonomy under some description and argue that moral obligation can be grounded in autonomy, the central elements of their accounts of autonomy are radically different. Indeed, Kant rejects each of the key inferences in Butler's account of moral obligation and autonomy. Where Kant argues that moral obligation must be grounded in the autonomy of reason, Butler argues that moral obligation can be grounded in the autonomy of human nature as a whole. And although both Kant and Butler both share some similarities with ancient Stoic moral theory, Butler's Stoic conception of virtue is unavailable to Kant.

2 KANT ON EMPIRICAL MORAL PHILOSOPHY

Even before delving into any of the details of Kant and Butler's accounts of autonomy and moral obligation, we have good reason to be wary of assimilating them. Most importantly, their characterizations of their methods for demonstrating moral obligation contradict one another. In an often-cited passage from the preface of his *Fifteen Sermons*, Butler claims that the obligation to virtue can be demonstrated in two ways. First, one can begin by "inquiring into the abstract relations of things" (P: 12). Butler is referring here to the method adopted by philosophers such as Samuel Clarke, who deployed a priori arguments aimed at demonstrating the obligation to virtue. This method, Butler claims, demonstrates that "vice is contrary to the nature of reasons and things" (P: 12). Butler himself does not adopt this method of demonstrating the obligation to virtue, but importantly he avoids doing so not because it leads to a different conclusion, but instead because his aims are largely homiletic, and the second method is especially "adapted to satisfy a fair mind; and is more easily applicable to the several particular relations and circumstances in life" (P: 12). This second method begins with "a matter of fact" about "the particular nature of man" (P: 12). Here, the point is to show that human moral

psychology itself obligates us to act virtuously. The details of Butler's argument will be the topic of section two, but for now it is enough to note that he thinks that if he can show that what he calls the "whole system of human nature" is characterized by the authority of conscience, then it follows that we are obligated to follow conscience. That is, Butler's account of autonomy, in its most general formulation, amounts to an inference from human nature to moral obligation.

Kant repeatedly and emphatically rejects grounding moral obligation in *any* facts about the particularities of human nature. For example, in the preface to his *Groundwork*, he famously claims that "the ground of obligation must not be sought in the nature of the human being or in the circumstances of the world in which he is placed, but a priori in pure reason" (4:389; c.f. 4:406; 4:425).⁵ This claim expresses Kant's rejection of a wide range of approaches to moral philosophy, including those of Butler, the British and Scottish sentimentalists, many of the ancient *eudaimonists*, and, to some extent, even some post-Kantian German Idealists, such as Hegel. The approaches Kant rejects here share the common conviction that empirical facts about nature, human psychology, or human society are relevant for grounding moral obligation.

Kant's rejection of empirical approaches to moral philosophy rests on his understanding of the requirements for justifying what he regards as "the common idea of duty and of moral laws," according to which moral laws establish obligations, and obligations are universally and necessarily binding (4:389). Because Kant believes that empirical observations cannot figure into any good argument for the universality and necessity of obligations, he rejects all approaches to moral philosophy that begin with empirical observations and opts instead for *a priori* argumentation. The pertinent question is then the following: why does Kant believe that empirical observations cannot figure into any good argument for moral obligation?

⁵ References to Kant's works in what follows are to the standard pagination of the *Gesammelte Schriften* and to the Cambridge edition English translations of Kant's collected works.

In the preface to the second *Critique*, Kant couches his answer to this question in terms of his broader response to Hume's empiricism (5: 12-3). Kant's most important and well-known engagement with Hume takes place within theoretical philosophy and is about the concept of causality. For Hume, empirical observation often shows us that an event A has always preceded an event B in time. However, when "we look about us towards external objects, and consider the operation of causes, we are never able...to discover any power or necessary connection; any quality which binds the effect to the cause, and renders the one an infallible consequence of the other" (*ECHU*: VII). That is, the observation that event A has always followed B cannot establish a strictly necessary connection between A and B. Hume thus concludes that "there is not, in any single, particular instance of cause and effect, any thing which can suggest the idea of power or necessary connection" (*ECHU*: VII). The concept of causality, however, is traditionally understood as claiming that if event A causes event B, then B *necessarily* follows A. As is well known, Kant accepts Hume's arguments that *empirical observation* cannot ground the claim to the necessary connection between two events made by the concept of causality but rejects Hume's conclusion that that this claim to necessity is spurious.

The basic Humean insight that awakens Kant from his "dogmatic slumber" is not limited to causal claims. On the contrary, Kant thinks that Hume's arguments against the necessity of causal connections relies on a more fundamental insight regarding what empirical claims are *about*, and this insight has ramifications throughout Kant's work. For both Hume and Kant, empirical observations are about particular circumstances and they are thus incapable of grounding any universal truths. Moreover, because claims about particular circumstances can only be contingently true, empirical observation cannot yield necessary truths. Hume's insight about empirical claims leads Kant to claim in the introduction to the second *Critique* that "[i]t is

an outright contradiction to want to extract necessity from an empirical proposition (*ex pumice aquam*) and to give a judgment, along with necessity, true universality” (5: 12).

Therefore, empirical methods in practical philosophy can yield “a practical rule, but never a moral law” (4: 389).⁶ Practical rules are conditional recommendations for action (5: 19-21). Using a favorite example of Kant’s, one could recommend as a practical rule that, if I want to be trusted, I ought to be honest. Because practical rules introduce a condition for the recommendation of some action (in our example this condition is wanting to be trusted), their normative force is contingent on the particular circumstances embodied in their condition (in our example these circumstances are those in which I want to be trusted). Empirical observation can establish practical rules precisely because they are particular and contingent in this sense—they only have normative force in particular circumstances, and their normative force is thus contingent on those circumstances obtaining. Moral laws, on the other hand, command unconditionally. For example, someone might simply claim that I ought to be honest. Moral laws state that I am obligated to act in a particular way regardless of whatever preferences I might have or what circumstances I might be placed in. In this sense, they are universally and necessarily valid. Kant assumes as a starting point in his practical philosophy that the common conception of morality involves moral laws of this sort; his arguments are aimed at showing that this common conception of morality is correct, that moral laws are in fact universally and necessarily valid.

Unlike Hume, Kant is not led by their common understanding of the character of empirical observation to abandon claims of absolute universality and necessity. He instead

⁶ It should not be surprising, given Hume and Kant’s shared understanding of empirical observation and Hume’s empirical method, that Hume’s moral philosophy establishes *utility* as the principle that accounts for our approval or disapproval of various character traits. In other words, for Hume, our approval or disapproval of character traits is precisely contingent on whether or not we desire the states of affairs that those character traits tend to produce.

attempts to find a way of grounding such claims that does not rely on empirical observation. In theoretical philosophy he thus attempts to show that empirical experience of the natural world is possible only if certain transcendental concepts, such as causality, structure that experience. And in his practical philosophy, Kant attempts to establish the obligation to act out of respect for the moral law by appealing only to pure reason.

However, this final claim must be qualified in order to remain coherent. Obligations are expressed in ‘ought’ claims, which Kant refers to as ‘imperatives’.⁷ Imperatives claim that performing some action is rationally required. For example, I might be presented with the following imperative: ‘you ought to pay back a friend from whom you have borrowed money’. For Kant, such represents a law—that borrowed money should be paid back—that I am obligated to obey. Importantly representing this law through an ‘ought’ claim only makes sense for a being who can do otherwise, who can act contrary to the law. That is, only beings that can act contrary to the law represent laws *as* imperatives. For this reason, “no imperatives hold for the *divine* will and in general for the *holy* will: the ‘ought’ is out of place here, because volition is of itself necessarily in accord with the law” (4: 414; c.f. 5: 32). What Kant calls a holy will, or a perfectly rational being, would be commanded by the same law but would necessarily act in accordance with that command. The command would then not be represented as an imperative. Only *imperfectly* rational agents such as ourselves represent objective laws of reason as imperatives. The domain of the moral law extends over all rational beings, but only imperfectly rational agents bear the relation of *obligation* to the moral law (5:32; 4: 413; 4: 439). Identifying the concept of obligation then requires some information about the practical psychology of finite

⁷ “The representation of an objective principle, insofar as it is necessitating for a will, is called a command (of reason), and the formula of the command is called an **imperative**. All imperatives are expressed by an *ought* and indicate by this the relation of an objective law of reason to a will that by its subjective constitution is not necessarily determined by it” (4: 413).

rational agents. In particular, the practical psychology of morally obligated agents must include the capacity to be motivated to act by the representation of the moral law and the capacity to act contrary to the moral law. The existence of beings with such a practical psychology is not derivable from pure reason, and so part of the preconditions of identifying the concept of obligation must be given empirically.

Although the concept of obligation applies only to finite rational agents and requires some information about their practical psychology, for Kant, the *ground* of obligation is nevertheless the part of that practical psychology that Kant calls ‘pure reason.’ Kant’s precise arguments for moral obligation shift from the *Groundwork* to the second *Critique*. But he remains convinced throughout his mature period that the moral law is derivable from pure reason alone, without reference to any empirical information, and that finite rational agents bear the relation of obligation to the moral law.

We are now in a position to see, if only in very general and rough terms, how Kant’s arguments against empirical methodology in moral theory apply to Butler. Butler does not ground his account of moral obligation in anything like Kant’s pure reason. He grounds it instead in a “matter of fact” about “the particular nature of man” (P: 12). Kant’s claim that moral obligation is grounded in the autonomy of pure reason is specifically intended to *avoid* relying on any facts about the particularities of human nature. Even before looking at the details of Butler or Kant’s arguments about autonomy and moral obligation we can then see that, because Kant rejects Butler’s inference from human nature to moral obligation, what Butler means by man being “a law to himself” is unlikely have much in common with Kantian autonomy (P: 29; 2: 8; 3: 5). As will become clear in the next three sections, however, such general observations about Kant and Butler’s methods do not establish the most significant differences between

Kantian and Butlerian autonomy. Indeed, Kant and Butler's projects are different in a way that makes their programmatic statements about methodology somewhat misleading when used to track the more fine-grained aspects of their respective conceptions of, and arguments about, autonomy and moral obligation. Seeing these differences then requires delving into some of the details of these arguments.

3 BUTLER ON OBLIGATION AND HUMAN NATURE

For Butler, moral obligation is grounded in the fact that human nature is a law to itself (P: 29; 2: 8; 3: 5). Understanding this claim requires providing answers to three questions. First, what, precisely, does Butler mean by "obligation?" Second, what is the conception of human nature that grounds obligation for Butler? Third, how does this conception of human nature ground obligation?

Butler's primary argumentative opponents in the first three sermons are those who recognize what courses of action conscience recommends, and that virtue requires following conscience, but nonetheless deny that there is any *obligation* to follow conscience. Conscience often seems to recommend imprudent courses of action, ones that conflict with our interests as individuals.⁸ For example, conscience might recommend that I put my own life at risk for the sake of saving the life of a friend. Moreover, I might recognize that risking my own life in this circumstance is genuinely the virtuous course of action. In such a case, I might still fairly ask myself whether the fact that conscience recommends that course of action provides me with a decisive or overriding reason to pursue that course of action. That is, I might find myself thinking that, although conscience claims that I *should* risk my life to save my friend, the fact

⁸ As we will see, Butler's ultimate position is that such conflicts are, at least, much less frequent than is often supposed, but he grants his opponents this point for the sake of showing that, even supposing such conflicts are frequent, we are still obligated to act virtuously.

that my own interests would be radically compromised by doing so means that I am not under any strict *obligation* to do so.

Butler thinks that he can show that we do have overriding reason to act virtuously, that we are obligated to virtue, even when the demands of virtue recommend a seemingly imprudent course of action. His argument relies on his account of human nature, which in turn is characterized by the authority of conscience (P:14; 2:1 3:1). Getting clear about human nature and obligation in Butler then requires at least a rough sketch of his argument(s) for the authority of conscience. The precise shape and aim Butler's argument or arguments for the authority of conscience is the subject of scholarly disagreement, most of which has no bearing on my argument here.⁹ For ease of exposition, and because I think it gets many of the central elements of Butler's argument right, I adopt the account of the authority of conscience developed by Brownsey. His interpretation of Butler's argument for the authority of conscience is straightforward. First, conscience *claims* to be authoritative (Brownsey 1995: 72-4). That is, Butler thinks that we experience conscience as authoritative in the sense that we experience conscience as recommending actions that ought to be performed whether or not they conflict with the recommendations of other parts of human nature. Butler expresses this point in several ways, claiming that conscience "magisterially asserts itself," that it is "manifestly" superior to all other principles in human nature, and that it is "a constituent part of the idea" of conscience that

⁹ Sturgeon argues for the controversial claim that, because conscience's approval and disapproval is grounded in judgments about naturalness and unnaturalness, Butler's argument for the authority of conscience entails that the authority of conscience is superfluous for Butler. Most commenters since Sturgeon argue against this view (Sturgeon: 1976). Darwall argues that conscience is self-authorizing in a sense that hews closely to Kantian autonomy (Darwall: 1991) Penelhum reads Butler as claiming that conscience's authority is derived from its access to independently existing facts about rightness and wrongness, bringing Butler into close contact with the rationalist tradition (Penelhum: 1985). Wedgewood in effect argues that Butler argues for the authority of conscience by enumerating some facts about the features of conscience—that it oversees other principles in human nature, that it judges about rightness and wrongness, and that we are disposed to moved to action by conscience (Wedgewood: 2007). I sketch Brownsey's interpretation below.

it “preside and govern” over the rest of human nature (2: 8; P: 24; 2: 14). For Butler, conscience is essentially characterized by its claim to authority; failing to recognize that conscience claims supreme authority is a failure to understand what conscience is.

However, as Brownsey points out, claiming to be authoritative is not the same as having legitimate authority. In order to establish the latter claim, Butler relies on the assumption that conscience must have *some* role to play in human life. Because (1) the only role conscience could have in human life must be commensurate with the kind of thing that it is, and (2) conscience is characterized by its claim to supreme authority, (3) conscience must *in fact have* supreme authority (Brownsey 1995: 74). According to Brownsey, by its nature conscience *claims* authority, and the consideration of the role that this claim to authority could have in human life shows that its claim is *legitimate*, that conscience has genuine authority.

Butler’s argument for the authority of conscience thus relies on the prior conviction that human nature is systematic. Butler himself never uses the term, but there is broad scholarly agreement Butlerian systems are *teleologically* structured. Butlerian human nature is then perhaps best understood in terms of a broader debate about the scope of mechanistic modes of explanation.¹⁰ By the time Butler was writing, the kind of mechanistic world-view advocated by Bacon, Descartes, and Newton had already been developed in theories of human nature by figures such as Hobbes and Mandeville. For both Hobbes and Mandeville, human nature is a mere compound or aggregate of various parts. Each part of human nature governs the behavior of an individual in turn as its strength supersedes the strength of the others. Mandeville describes the process by which the strength of one part of human nature supersedes the strength of another

¹⁰ It is worth noting that, although Kant does not share Butler’s precise notion of systems, and, as I argue below, also does not understand human nature teleologically, Kant’s ‘Critique of Teleological Judgment’ is an important text in this debate. Both Kant and Butler are (rightly, I think) concerned with showing that mechanical-causal explanations cannot adequately account for all phenomenon.

in terms of one part being “provoked,” and governing a person “whether he will or no,” indicating that he understands human nature in mechanical-causal terms (36). Hobbes makes this point especially clear by classifying the study of human nature as a subset of physics, which concerns the “invisible motions of the internal parts of body” (I; ix; 5).

Both Hobbes and Mandeville make it clear that their conceptions of human nature do not provide any standards for moral assessment of behavior. Hobbes famously claims that various human desires and passions, and the actions that follow from them, are “in themselves no sin,” and are only right or wrong in relation to a conventional law. Moreover, without political authority, humans are naturally in a state of war, where “notions of right and wrong, justice and injustice, have...no place” (I; xiii; 8-13). Mandeville expresses a similar point with his usual dramatic flair, claiming that “[t]here is no merit in saving an innocent babe ready to drop into the fire” because “whatsoever benefit the infant received, we only obliged ourselves” (43). Mandeville, unlike Hobbes, is not concerned with establishing the legitimacy of political authority, but his basic point is the same—moral judgments cannot be grounded in human nature because no part of human nature has a legitimate claim to moral authority.

The lack of a principle in human nature that has the authority to make moral judgments entails that all behavior must be judged as equally appropriate or proper to human nature as a whole, and this is, for Butler, their key flaw. Butlerian teleology aims to ameliorate this flaw. Butler uses the example of a watch to illustrate the sense in which he regards human nature as a teleological system. Butler notes that understanding the constitution of the various parts of a watch is insufficient for understanding the “idea of a watch.” This latter task requires forming “a notion of the relations which [its] several parts have to each other—all conducive in the respective ways to this purpose, shewing the hour of the day” (P: 14). A watch functions

properly to the extent that its parts individually, and in their relations to one another, contribute to the function of the whole—namely telling time. From these brief remarks we can draw out the three aspects of Butler’s account of teleological systems that are important for present purposes: (1) there are proper and improper ways in which it can function, (2) the whole system is disposed towards functioning properly, and (3) the functions of the parts of the system are related to one another in a way that promotes the proper functioning of the whole system.

The idea that there are proper and improper ways that watches can function, and that the parts of the watch considered individually and the whole watch are disposed towards functioning properly (provided that the watch is reasonably well made), is surely uncontroversial. However, one might fairly think that some relevant disanalogies arise when applying these aspects of Butlerian teleology to human nature. The purpose of a watch is to tell time, and the whole watch and its parts function properly precisely to the extent that they serve this purpose. But discerning a similar criterion for determining whether or not the system of human nature is functioning properly is more difficult.

Butler makes it clear that he thinks the authority of conscience introduces the criterion that allows us to distinguish between proper and improper functioning of the system of human nature.¹¹ His example of parricide can help illustrate his position. As we have seen, Butler’s conception of human nature is opposed to those that understand the relations between that parts of human nature only in terms of strength, and not in terms of legitimate authority. Butler claims

¹¹ The claim that the authority of conscience provides the criterion for distinguishing between the naturalness and unnaturalness of action should not be read as entailing Sturgeon’s “Full Naturalistic Thesis,” which states that “conscience never favours or opposes any action, except on the grounds which include its naturalness or unnaturalness” (Sturgeon 1976: 328). The authority carried by judgments of conscience about the rightness or wrongness of actions considered ‘in themselves’ is what provides a criterion for distinguishing between naturalness and unnaturalness of actions. These judgments, for this reason, must be independent of and prior to judgments of naturalness.

that, if such conceptions of human nature are correct, then “we can make no distinction between these two actions [the murder of one’s father and an action that accords with one’s nature], considered as the actions of such a creature; but in our coolest hours must approve or disapprove of them equally” (2: 17). Butler regards this consequence as absurd—we are plainly able, he thinks, to correctly regard some actions, such as the murder of one’s father, as contrary to our nature.

We are able to do so on Butler’s view because the disapproval of conscience at the murder of one’s father carries legitimate authority, and, most importantly, this authority characterizes the whole system of human nature. Like the way in which the whole constitution of a watch is determined by the function of telling time, it “is from considering the relations which the several appetites and passions in the inward frame have to each other, and above all the supremacy of...conscience, that we get the idea of the system or constitution of human nature” (P:14).¹² Because the constitution of human nature is characterized by the authority of conscience, and conscience recommends virtue, human nature is “adapted to virtue” (P: 14). This final claim is best understood as a *description* of human nature.¹³ For Butler, descriptions of systems refer to their functions. For example, any adequate description of a watch refers to the fact that it is for telling time. This also introduces a normative component—a watch is functioning properly to the extent that it can be used to tell time. Analogously, any adequate

¹² C.f. (P: 14) Once we have in view this idea of the whole system of human nature, “from the idea itself it will appear, that this is our nature, *i.e.* constitution is adapted to virtue, as from the idea of a watch it appears that its nature, *i.e.* constitution or system is adapted to measure time”.

¹³ One might object here that Butler is guilty of committing that we might call the “inverse naturalistic fallacy.” That is, he might be guilty of attempting to derive an “is” from an “ought.” This objection would rely on the mistake that Butler’s claim that the whole of human nature is adapted to virtue makes a direct claim about how the parts of human nature *in fact* relate to one another. Butler is certainly not claiming that humans usually act according to their conscience, and he makes it clear that he thinks that systems are often “apt to be out of order” (Preface: 16). A system is out of order to the extent that its parts do not in fact relate to one another in the way demanded by the character of that system. The important thing to note is that, for Butler, acting contrary to conscience is a *perversion* of the system of human nature.

description of human nature must refer to the fact that its parts work together to promote virtue, and that, therefore, human nature is functioning properly to the extent that virtue is promoted.

This description of human nature, for Butler, entails that man is “a law to himself,” and that we are thus “obligated to virtue.” He seems to think that the obligation to virtue follows in some obvious sense from his account of human nature, going so far as to claim that the question of whether we are obligated to follow “the rule of right within” ourselves “carries its own answer along with it” (3: 5). Butlerian human nature is then autonomous in the sense that its prescriptions about the naturalness or unnaturalness of actions have the force of a law. This law establishes a strict obligation—that is, for Butler, we *always* have overriding reason to act in whatever way is natural for us. Butler’s argument for moral obligation thus states that we are obligated to do the right thing, as judged by conscience, because doing otherwise would be constitute a violation of the law provided by human nature.

We can then answer the three questions with which this section began: Butler thinks that we are morally obligated to behave in some way if our nature provides us with overriding reason to behave in that way. For Butler, conscience is the part of human nature that judges about whether or not actions and character traits are virtuous or unvirtuous, and its judgments carry supreme normative authority. The supreme normative authority of conscience yields a claim about human nature as a whole—human nature is adapted to virtue in the sense that human beings obey the law of our nature precisely to the extent that we follow conscience and thereby behave virtuously. Because human nature as a whole prescribes virtuous life as a law—and is autonomous in precisely this sense—human beings are obligated to virtue.

We saw in section one that Kant regards any inference from the particular constitution of human nature to moral obligation as invalid. Even if Kant shared Butler’s teleological conception

of human nature he would regard Butler's account of moral obligation as unsatisfactory. In this sense, Kant would reject the inference that Butler makes from the whole of human nature to moral obligation. However, the full scope of the differences between their conceptions of autonomy is not intelligible in terms of this disagreement alone. Kant's rejection of inferences from the particularities of human nature to moral obligation is, as we have seen, rooted in his concern with empirical methodology in moral philosophy. But it is far from clear that Butler's methodology is empirical in a sense that meaningfully distinguishes it from Kant's. Butler may be interested in justifying claims about moral obligation on the basis of some purported facts about human nature, but these 'facts' share a good deal in common with the starting points of Kant's arguments.

Conscience's 'magisterial assertion' of its own authority may look like a contingent empirical experience, but there is little difference between it and Kant's "fact of reason," aside from Kant's bare insistence that the latter "is not an empirical fact but the sole fact of pure reason" (5: 31). Butler does not share, and likely would not be especially concerned about, Kant's conception of pure reason, and thus does not have this characterization of the magisterial assertion of the authority of conscience available to him. However, both the authority of conscience and the fact of reason are fundamental in the sense that they cannot be reasoned out from antecedent data, and for this reason they are both *starting points* for argumentation. Moreover, it is unclear what empirical experience could possibly underwrite Butler's claim that human nature is a teleological system. The fact that every system is "apt to be out of order" entails that observing the *actual* behavior of human beings cannot establish claims about what counts as the proper function of human nature, and this is part of Butler's reason for insisting that "what actually happens is of no concern" for his claims about the whole system of human nature.

Merely relying on Kant's arguments against empirical methodology in moral philosophy for understanding the differences between Kantian and Butlerian autonomy is thus insufficient. The primary Kantian reason for rejecting Butlerian autonomy is that Kant does not think that any claims about the particular and contingent constitution of human nature can ground moral obligation. Understanding the Kantian perspective then requires examining the sense in which, for Kant, the ground of moral obligation is independent from the particularities and contingencies of human nature.

4 KANT ON REASON AND HUMAN NATURE

For Kant, moral obligation is grounded in the autonomy of what, by the *Critique of the Power of Judgment*, he comes to call moral-practical reason, which he distinguishes from technical-practical reason and theoretical reason. I will simply use the term 'reason' to refer to moral-practical reason. As was noted above, it is important for Kant's project of grounding moral obligation in the autonomy of reason that reason is independent from human nature. However, as will become clear, reason also must, in some sense, be a *part* of human nature. Understanding precisely what it means for Kant to affirm both of these claims, and what his affirmation of both of these claims means for the differences between Kant and Butler's conceptions of autonomy, is the task of this section.

As we have seen, for Kant, the problem with grounding moral obligation in human nature is that moral obligation is universal and necessary, while human nature is particular and contingent. Nothing *guarantees* that human nature is constituted in whatever way it is, and the constitution of human nature is, of course, specific to humans. The domain of the moral law, however, extends over all rational entities regardless of however else they might be constituted, and acting according to its demands is required of all rational entities—this is the sense in which

the moral law is universal and necessary. Because nothing universal and necessary can be grounded in anything particular and contingent, the moral law cannot be grounded in human nature. Kant's task is then to show—in a way that avoids relying on information that is specific to human beings—that there is a moral law. He repeatedly insists that the only way to accomplish this task is to rely only on 'pure reason.'

For this method to be a sensible way of avoiding the problem of grounding a universal and necessary claim in particular and contingent truths about human nature, reason must be independent of human nature in a sense yet to be specified. Herein lies the significance of Kant's repeated claims that the domain of the moral law extends not only over humans but over all possible rational beings, including holy wills and God, and that "*only because of this* [can it] be also a law for all human wills" (4:425; c.f. 5: 32). A condition of the genuine moral authority of the moral law, for Kant, is that it is grounded not in "a special tendency that would be peculiar to human reason," but in reason construed independently from any particularities of human psychology (4: 425). That is, reason in Kant's sense is not constituted by rules of merely *human* thought, but by rules of *any and all* thinking properly so called.¹⁴

Although Kantian reason cannot be understood as a "special property of human nature," it must in some sense be a part of human nature. Kant frequently characterizes us as finite rational beings, and his argument that human beings must be afforded moral respect hinges on the claim that we are persons, distinguished from mere things by having a rational faculty, and are therefore "by [our] nature" ends-in-ourselves.¹⁵ Moreover, were Kantian reason not a part of

¹⁴ This is, I take it, part of the force of the previously cited passage, and many of those that claim that the domain of reason's moral law extends over all rational beings.

¹⁵ "Beings the existence of which rests not on our will but on nature, if there are beings without reason, still only have a relative worth, as means, and are therefore called *things*, whereas rational beings are called *person* because their nature already marks them out as an end in itself" (4: 428)

human nature, this would seem to commit him to the unfavorable position of offering a moral theory that has no genuine bearing on human beings. In an especially helpful passage in the *Groundwork*, Kant sums up the aspects of his account of the relationship between reason and human nature with which I am presently concerned. He claims that the “human being is bound to laws by his duty...he is subject *only to laws given by himself but still universal* and...he is bound to act in conformity with his own will, which, however...is a will giving universal law” (4: 432). Kant’s claim that we human beings *give ourselves* the moral law makes sense only on the condition that reason is an essential characteristic of humanity, and is in this sense a part of human nature. When we act morally, we act according to *our own* will, which is characterized by its ability to prescribe a genuinely universal law. In this sense, reason is a part of human nature. The idea that Kantian reason is independent of human nature ought to be understood as stating that reason is not dependent on its instantiation in human nature, not as stating that it is not instantiated in human nature at all.¹⁶ The human mind, for Kant, is essentially characterized by rationality, but reason is not essentially characterized by its presence in the human mind. Its claims can therefore be universal and necessary in Kant’s strong sense of those terms, and still bear meaningfully on *human* morality.

¹⁶ Some scholars have located a tension here between the claims that obligation must be grounded in reason, not in human nature, and that reason is a part of human nature. They have often attempted to resolve this tension by claiming that the term ‘human nature’ in each claim has distinct referents. In the first claim, ‘human nature’ refers to some empirically observable facts about human behaviors, tendencies, and preferences of the sort that Kant investigates in his *Anthropology*. In the second claim, the term refers to our *transcendental* nature, which is not amenable to empirical observation, and can only be investigated a priori. See (Frierson: 2013) for an admirably clear exposition of this distinction. On this line of thought, Kant rejects grounding obligation in human nature construed empirically, and he instead opts for grounding moral obligation in human nature construed transcendently. Although it is surely fair to use the term in both ways in reference to Kant’s conception of the human being, using this distinction does little to settle many of the most pressing questions about the relationship between reason and human nature, and largely serves to obfuscate the force of Kant’s insistence that obligation cannot be grounded in human nature. His arguments apply equally to any facts about human nature construed transcendently as they do to facts about human nature construed empirically.

Without getting into the details of Kant's arguments for the autonomy of reason, we can already see what the broad outlines of his account of how autonomy grounds moral obligation will look like. Although, as Kant famously puts it in the *Groundwork*, "there is nothing in heaven or on earth" on which moral philosophy is based, it must nevertheless "manifest its purity as sustainer of its own laws" (4: 425). Reason, for Kant, is self-justifying in its claims and is the source of its own laws, and this is the sense in which it is autonomous. Reason's moral law carries its own authority with it, and rational beings are therefore *obligated* to uphold that law. Kant's account of obligation and autonomy then differs substantially from Butler's. Where, for Butler, the *whole* of human nature provides itself with a law that establishes the obligation to follow conscience, for Kant, the *part* of human nature called 'reason' provides human beings with a law that we are obligated to uphold.

However, this still does not exhaust the differences between Kantian and Butlerian autonomy. For Kant, moral obligation is sufficiently established by the authority of reason, without any arguments about that authority characterizing the whole of human nature. Kant might then be in a position to claim that Butlerian conscience is, although Butler did not fully realize it, self-authorizing and that his conception of the whole of human nature was unnecessary for establishing moral obligation.¹⁷ That is, if Kant, like Butler, had a conception of the whole of human nature rooted in the authority of reason, then even though Kant would have good reason to regard Butler's argument about the whole of human nature *superfluous* for grounding moral obligation, he would not have good reason for *rejecting* it. As we will see, however, Kant does have good reasons for rejecting Butler's conception of the whole of human nature, and all of Butler's conception of autonomy along with it.

¹⁷ Darwall pursues an interpretation of Butler's arguments for the authority of conscience according to which Butler did think that conscience is self-authorizing in a Kantian sense. (Darwall 1991)

5 KANT AND BUTLER ON HUMAN NATURE

We saw in section two that Butler's conception of autonomy contains three key inferences. Butler infers from the part of human nature called 'conscience' to a claim about the whole of human nature—that it is adapted to virtue. The second inference, which, for Butler, validates the first, is from norm to fact. The *normative authority* of the part of human nature that Butler calls conscience yields the aforementioned *descriptive fact* about the whole of human nature, that it is adapted to virtue. This fact about human nature is descriptive the way that one describes what a watch is by saying it is for telling time—that is, saying that human nature is adapted to virtue is at once a way of saying what human nature is and prescribing its proper function. The third inference, which, as we saw in section one, Kant rejects, is from a particular fact about the whole of human nature—that it is adapted to virtue—to moral obligation. In this section, I argue that Kant also rejects the first two inferences, and that Kant and Butler therefore do not share *any* of the key components of their conceptions of autonomy. Their disagreement about inferring from the authority of a part of human nature to a description of the whole of human nature can be shown most easily by noting a direct consequence of Butler's view that the whole system of human nature is adapted to virtue. Namely, Butler's conception of human nature entails that all the parts of human nature are best understood in terms of their tendency to promote virtue. This consequence is absent in Kant.

Butler's position that all parts of human nature are best understood in terms of their tendency to promote virtue is implied by the third aspect of Butler's conception of systems articulated above, according to which the functions of the parts of a system relate to one another in a way that promotes the proper functioning of the whole system. For Butler, the proper functioning of the whole system of human nature is prescribed by conscience. For this reason,

the proper functioning of *each part* is prescribed by conscience, and each part of human nature in itself tends towards virtue. Butler insists in several contexts that other parts of human nature have fulfill their genuine function to the extent that that promote virtue and happiness and violate that function the extent that they promote viciousness and misery.¹⁸ Many of Butler's sermons are dedicated to demonstrating precisely this point, and operate on the basic principle that no part of human nature that "God hath endued us with can be in itself evil." It follows for Butler that when evil effects follow from some part of human nature they become "quite another thing from what [they were] originally in our nature" (8: 3). Thus, when resentment is directed against its proper object, "vice and wickedness," it is "one of the common bonds by which society is held together" (8: 7). Even genuine self-love is "one chief security of our right behaviour towards society" and coincides perfectly with benevolence (1: 6; c.f. P: 28; P: 39 P: 41; 5: 1). The kind of bad self-partiality that is the root of so much vice can then rightly be called "false self-love" (9: 20). That is, when parts of human nature such as resentment and self-love lead to vicious behavior, for Butler this is not merely a failure of conscience, but also a failure of resentment and self-love themselves.

One must be careful to avoid misunderstanding Butler's position on these points. For Butler, although the parts of human nature must be understood in terms of their tendency to promote the ends of conscience, they are not *reducible* to conscience. Parts of human nature such as self-love and resentment are not "conscience in disguise." Butler insists that "Everything is what it is, and not another thing" (P: 39). Thinking otherwise would involve a mistake similar to the kind Butler sees in proponents of the "selfish hypothesis," such as Mandeville and Hobbes, who believe that principles such as benevolence and compassion are reducible to self-love.

¹⁸ As we will see, it is important for Butler that happiness, in general, coincides with virtue, and that misery, in general, coincides with vice.

Butler's conception of human nature consists in a system of several *distinct* parts that, when governed by their proper function, work in harmony with one another to promote virtue. For example, genuine self-love, for Butler, recommends actions that happen to benefit others. But this does not mean that genuine self-love is *identical* to benevolence. On the contrary, I may often behave in a way that benefits others largely because it promotes my own happiness, and it may promote my own happiness because I have a genuine concern for others. In such cases, benevolence and self-love work in harmony with one another to promote virtue.

Kant, on the other hand, usually views elements of human nature such as passions and affects as opposed to, or as obstacles in the way of, fulfilling the demands of moral reason. Affects, for Kant, "belong to *feeling* insofar as, preceding reflection, it makes this impossible or more difficult" (6: 407). Because affects precede reflection, they frustrate our attempts to act out of respect for duty. Passions, unlike affects, permit reflection, but are no less opposed to the demands of moral reason. On the contrary, for Kant, because passions permit reflection, they can become "rooted deeply, and so take up what is evil (as something premeditated) into [their] maxim" (6: 408). That is, passions can become the cause of genuine, reflectively endorsed, evil. The important thing to note here is that these elements of human nature are not appropriately understood in terms of their tendency to promote the ends of moral reason. Passions and affects are not functioning improperly when one acts from them and against moral reason. For example, if my immediate, affective response to an insult leads me to attack my detractor, this, for Kant, is not a failure of my *affect* to be directed in the right way to the right object. It is a failure of *reason* to regulate my affect. Or if my passionate disdain for a peer leads me to try to ruin his career, once again this is not a failure of my *passion* but a failure of my *reason*. Even if it turns out to be true that Kantian virtue requires properly cultivating all parts of human nature, this

requirement is quite different from the Butlerian claim that all parts of human nature *in themselves* tend towards virtue. This latter claim is the issue here, and Kant certainly does not agree with it.

Kant's understanding of passions and affects indicates that he, unlike Butler, does not think that his commitment to the authority of moral reason entails any descriptive claim about human nature as a whole. Kant's disagreement with Butler can be understood in terms of Kant's opposition to Butler's systematicity thesis about human nature. As we have seen, Butler's argument that human nature as a whole is conducive to virtue relies on two key premises. First, conscience is supremely authoritative. Second, human nature is a teleological system the function of which is prescribed by conscience. Kant and Butler, roughly speaking, agree on the first premise. The second premise, however, entails a position about elements in human nature other than moral reason that Kant explicitly opposes. He does not agree with Butler that human nature is a teleological system, and Butler's position that human nature as a whole is adapted to virtue is thus unavailable to him.

Further differences between Kant and Butler's conceptions of autonomy are intelligible in terms of their disagreement about the validity of inferring from the normative authority of a part of human nature to a description of the whole of human nature. For Butler, human nature is a teleological system whose function is expressed by authority of conscience. Because teleological systems, for Butler, are conducive to their function, and conscience recommends virtue, human nature is appropriately described in terms of its conduciveness to virtue. Butler's description of the whole of human nature as a teleological system whose proper function is prescribed by conscience allows him to yield a moral claim from it. The obligation to obey conscience is grounded in the fact that "man is *by his nature* a law to himself" [italics added] (3:

5). Kant's conception of human nature, as we have seen, does not include the aspects of Butler's teleological account of human nature that underwrite Butler's attempt to ground obligation in it. As is shown by Kant's understanding of parts of human nature as obstacles to morality, such as inclination and passion, Kantian human nature is not a system whose proper function is prescribed by moral reason. Kant then rejects all of the key aspects of Butlerian autonomy.

6 BUTLER ON STOIC VIRTUE

Questions about what kind of creatures we are, or of human nature broadly construed, lie at the heart of both Butler and Kant's philosophical projects. The above claims about human nature then ought to have far-reaching implications for what further positions are and are not available to each thinker. The rest of this thesis traces one such implication to a negative result. My argument is straightforward. Because Kant does not believe that human nature as a whole is conducive to virtue, his conception of virtue cannot consist in following human nature. Put differently, Butler's Stoic conception of virtue is unavailable to Kant. In order to show why Butlerian Stoic virtue is unavailable to Kant, it is useful to articulate some of the more important aspects of Butler's conception of virtue. For present purposes, two of Butler's theses regarding virtue, the first of which he shares with the ancient Stoics, are most important. First, virtue consists in following one's whole nature, and second, virtue tends towards happiness.

Butler's stated goal in his first three sermons is to provide an interpretation and defense of the "inward feeling" of the "ancient moralists" that "man is born to virtue, that it consists in following nature, and that vice is more contrary to this nature than tortures or death" (Preface: 13). As was noted above, scholars agree that the ancient moralists to whom Butler refers here are the Stoics. Long has argued that Butler follows the Stoics in distinguishing between three senses that the term "nature" can have in discourse about following one's nature (Long 2003: 387).

Butler articulates this distinction the most clearly in the second sermon. He writes, “By *nature* is often meant no more than some principle in man, without regard to either the kind or degree of it. Thus the passion of anger, and the affection of parents to their children, would be called equally *natural*” (2: 5). For Butler and the Stoics, satisfying an appetite, pursuing the object of a desire, and acting in any way that is conducive to the material well-being of ourselves are all natural for us. Because all action follows from some principle in human nature, all action counts, in this sense, as following one’s nature. However, “[*n*]ature is frequently spoken of as consisting in those passion which are strongest, and most influence actions” (2: 6). In this second sense of nature, because all action for Butler follows from the principle in human nature that is strongest at the moment of action, the principles that influence action the most frequently are those that are most properly called “natural.” Butler notes that this is the sense of nature that underwrites claims about the viciousness of human nature—because humans often behave viciously, we are naturally vicious.

For Butler and the Stoics, neither of these senses of the term ‘nature’ are relevant to the moral principle that one ought to follow one’s nature. The relevant sense of “nature” here is the one articulated above, according to which human nature is a system of interconnected parts that each participate in the function of the whole. Virtue, for Butler and the Stoics, consists in following one’s *whole* nature, and, as we have seen, for Butler one following one’s whole nature requires acting according to conscience. Thus, “[*m*]an may act according to that principle or inclination which for the present happens to be strongest, and yet act in a way disproportionate to, and violate his real proper nature” (2: 10).

The sense of nature operative in the moral imperative to follow one’s nature, which Butler refers to as variously as “real,” “proper,” and “whole,” underwrites the best argument

available to Butler for the claim that virtue tends towards happiness. He, however, seems to think that a simple empirical observation is sufficient for showing that virtue tends towards happiness—if you observe the lives of vicious people, you will see that they usually tend to be miserable. Conversely, observing virtuous people reveals that they tend to be happier. Moreover, he thinks that everyone recognizes these facts, even if they also tend to think that happiness would consist in material gratification if religion were out of the picture.¹⁹ From such observations, Butler concludes that the virtuous life leads us “to attain the greatest happiness...in the present world” (1: 15).

Even leaving aside complications about attributing misery or happiness to others, this empirical justification of the coincidence of happiness and virtue is easy to attack. It may be the case that some forms of vicious life, such as those involving excessive consumption of mind-altering substances, do often end in misery. But this, of course, does not entail that the corresponding virtue (in this case, temperance) will lead to happiness. We also might disagree with the observation itself. Vicious people, one might think, often enjoy life far more than virtuous people. If this is right, then Butler’s empirical justification for the coincidence of virtue and happiness fails. Finally, and perhaps most persuasively, one might reasonably think that whether or not a person will likely be happy living a vicious life, or miserable living a virtuous life, must always be answered on a case-by-case basis. People’s tastes differ. There might then not be any good reason to expect that virtue would tend towards happiness or vice towards misery.

¹⁹ “persons in the greatest affluence of fortune are no happier than such as have only a competency; that the cares and disappointments of ambition for the most part far exceed the satisfactions of it; as also the miserable intervals of intemperance and excess, and the many untimely deaths occasioned by a dissolute course of life: these things are all seen, acknowledged, by everyone acknowledged” (*Sermons* 1:14).

Such objections to Butler's empirical argument that virtue tends towards happiness rely on the lack of a conceptual link between virtue and happiness. Although, as far as I can tell, Butler never explicitly provides such a conceptual link, he has the argumentative resources to do so available to him. Butler can argue that the psychic harmony brought about by virtue makes it tend towards happiness, and the psychic disharmony brought about by vice makes it tend towards misery, and that we thus have good reason to *expect* that the virtuous life is the happy life.

This argument begins with Butler's systematic conception of human nature. As we have seen, the authority of conscience grounds the account of human nature operative in Butler's claim that virtue consists in following human nature. Conscience governs how the various parts of human nature ought to relate to one another. The parts of human nature, and human nature as a whole, are functioning properly to the extent that the demands of conscience are heeded. Thus, the whole system of human nature is *in proper order* to the extent that we act in accordance with conscience. Butler compares a disordered person, one who does not act in accordance with their conscience, to a disordered work of art (P: 14). If the various parts of a work of art fail to work together in order contribute to the purpose of that work, then the work is rendered disorderly. The success of the work then relies the participation of each part in promoting the function of the whole. The same is true, for Butler, of human nature, the only important difference being that "we are agents" and "our constitution is put in our own power. We are charged with it: and therefore are accountable for any disorder or violation of it" (P: 14).

In the case of the system of human nature, each part participates in the function of the whole in the sense that, without the contribution of each part of human nature, the whole cannot function properly. The clearest example of Butler's position on this point rests in his claim that,

although a perfect being would perhaps be free from natural human affections, if *we* did not have them, we would be “without a sufficient principle of action.” The natural affections, such as “hunger, thirst, and weariness” therefore “are placed within as a guard and further security,” without which “our life would be neglected” even though “reason would assure us that the recruits of food and sleep are necessary means of our preservation” (5: 3). As we have seen, the same is true for all other parts of human nature for Butler, even those such as resentment and self-love—each has some role that it plays in promoting the proper functioning of the whole system of human nature.

Because a system functions properly only when each part of the system contributes to the function of the whole, properly functioning systems are *harmonious*. Insofar as a person leads a virtuous life, the various parts of their constitution work in harmony to promote virtue. The vicious person, on the other hand, is not afforded this harmonious interaction of the parts of their constitution. The internal harmony brought about by virtuous action provides the conceptual link between virtue and happiness that Butler’s empirical observations lack. Virtue tends towards happiness because virtuous life is more harmonious and peaceful than vicious life. The vicious person is still subject to the pangs of conscience but refuses to heed them, and therefore is made unhappy. In his “Dissertation on the Nature of Virtue,” Butler expresses precisely this point, writing that, “if human creatures are endued with...a moral faculty” then “moral government must consist, in rendering them happy and unhappy...as they follow, neglect, or depart from, the moral rule of action interwoven in their nature” (*D*: 9).

Importantly, the conceptual link between virtue and happiness does not guarantee that the virtuous person will always be happier than the vicious person. Butler is perfectly sensitive to the fact that there do seem to be some happy vicious people, and some miserable virtuous people.

For Butler, unlike the Stoics, the concept of happiness has a positive component—happiness requires the enjoyment of some object. For this reason, psychic harmony cannot by itself guarantee happiness for Butler. However, the psychic harmony brought about by virtuous behavior does make it clear why Butler thinks that we ought to usually *expect* virtue to tend towards happiness. The virtuous person is spared from suffering the psychic disharmony that comes with vicious behavior, and, as a general rule, she will then tend to be happier than the vicious person. Virtue consists in following our nature, and by following our nature we are internally harmonious—following our conscience is pleasant because in doing so we are most fully ourselves and therefore free from internal tension.

7 KANT ON STOIC VIRTUE

Kantian virtue differs from Butlerian virtue in many important ways. In this section, I argue that Butler's most general formulation of his account of virtue—where virtue consists in following human nature—is unavailable to Kant. In order to do so, it will be helpful to highlight some of the most important aspects of Kant's own account of virtue. In particular, I argue that Kant's conceptions of rational fortitude and autocracy, and his polemic against habit, suggest that human virtue does not typically involve psychic harmony. This lack of psychic harmony in virtuous life ought to be expected given Kant's conception of human nature. Because moral reason does not adequately describe the whole of human nature, virtue cannot consist in following human nature, and does not result in psychic harmony.

In an especially clear passage, Kant characterizes virtue in terms of a particular kind of strength. He writes that virtue is “the moral strength of a *human being's* will in fulfilling his *duty*, a moral *constraint* through his own lawgiving reason, insofar as this constitutes an authority *executing* the law” (6:405). Kant often refers to this kind of moral strength as autocracy

or self-command (*Selbstbeherrschung*). The virtuous life then is one in which an individual successfully commands herself. The details of Kant's conception of self-command cannot be worked out here. For Kant, the "*self-command* [*Selbstbeherrschung*] of practical reason...involves consciousness of the *capacity* to master one's inclinations when they rebel against the law" (6: 383).²⁰ Following Wilson's account of self-command, I take the capacity that Kant refers to here to be ethical self-legislation, which "involves (a) putting forward a proposition that represents a course of action as required and (b) attaching the proposition to a motivation to do what the proposition says must be done" (Wilson 2015: 262; cf. 6: 218). In the case of ethical self-legislation, the relevant motivation is respect for duty. Duty, of course, is not the only motivation for human action—human beings have the *capacity* to attach the idea of duty to an action as a motivation, but often fail to adequately *exercise* this capacity. Self-command then involves the consciousness of the capacity of self-legislation, the capacity to adopt duty as a motivation for action when our rebellious inclinations recommend actions that violate duty. The moral strength that Kant equates with virtue in the above cited passage, self-command is characterized by successfully acting in accordance with consciousness of the capacity of self-legislation. Kantian virtue can then broadly be understood as the proficiency in exercising the capacity of self-legislation, and thereby successfully reigning in one's rebellious inclinations when needed.

That Kant's understanding of virtue is characterized largely by strength in the face of rebellious inclinations is telling. He often characterizes virtue by analogy with fortitude, claiming that virtue, like fortitude, involves "the capacity and resolve to withstand an unjust opponent"

²⁰ Cambridge has opted to translate "*Selbstbeherrschung*" as "autocracy," presumably to mirror the term "autonomy." I have instead opted translate "*Selbstbeherrschung*" as "self-command," largely because in the German these terms do not mirror one another in the same way.

(6:380). However, in the case of virtue that opponent is “*within us*” (6: 380). As we have seen, Kant thinks that the primary obstacles to the demands of reason are internal to human nature. This is one of the primary reasons why Kantian human nature as whole cannot be described in terms of reason. We ought, then, to read Kant here as emphasizing that virtue involves overcoming the elements within human nature that are opponents of, or obstacles standing in the way of, the demands of moral reason. As early as the first *Critique*, Kant writes that “obstacles *in human nature*...may hold us at a distance from [moral perfection]” (emphasis added A: 315; B: 372). Importantly, the strength to overcome these obstacles cannot involve simply restraining one inclination in favor of another. Virtue instead involves governing, or perhaps even conquering,²¹ all inclinations in accordance with a rational principle. For Kant, inclination ensures that “virtue can never settle down in peace and quiet with its maxims adopted once and for all, but, if it is not rising, is unavoidably sinking” (6: 409).

Kant also insists that virtue properly so called cannot be habitual. His reasons for doing so are wrapped up in the notion that morally worthy actions adopt respect for duty as their determining ground and are therefore characterized by freedom from inclination. When we perform an action out of habit, its determining ground is inclination, and therefore “if the practice of virtue were to become a habit the subject would suffer loss to that *freedom* in adopting his maxims which distinguishes an action done from duty” (6: 409). Kant’s worry about inclination is not only that it frequently recommends actions that violate duty, but that acting out of inclination, even when it is in accordance with duty, cannot be a part of a genuinely virtuous life. Kantian virtue is perhaps best characterized as a sort of constant rational vigilance, where

²¹ “Impulses of nature...involve *obstacles* within the human being’s mind to his fulfilment of duty and (sometimes powerful) forces opposing it, which he must judge that he is capable of resisting and conquering by reason” (6: 380).

any coincidence between one's inclinations and one's actions often largely accidental. Because Kant understands happiness simply as the "sum satisfaction of all inclinations," any coincidence between virtue and happiness is likewise largely accidental (4: 399).

Kant's conception of "human morality in its highest stage," which is "commonly personified poetically by the sage," then looks very little like Butler's Stoic ideal of virtue (6: 383). In the latter, each part of human nature fulfills its proper function by pursuing its proper objects, and each contributes to the harmonious order of the whole by doing so. In this sense, Butlerian Stoic virtue consists in following human nature, and Butlerian virtuous life, for this reason, tends towards happiness. Kantian virtue, however, consists in the proficient exercise of ethical self-legislation. When inclination rebels against the demands of reason, the virtuous person motivates herself to behave morally through the idea of duty. The Kantian sage never acts merely out of habit or inclination, but instead uses reason to *overcome* the other parts of human nature that motivate action—especially affects and passions. She does not, then *follow* her nature in Butler's sense, and, if she is happy, that is largely a lucky accident.

As I indicated above, ethical self-legislation refers to the capacity to act against inclination out of respect for duty, and proficiency in exercising this capacity is virtue. At this point, one might raise an objection to my claim that Butler's Stoic claim—that virtue consists in following human nature—is unavailable to Kant. Throughout much of the history of western philosophy, following human nature has been understood as exercising 'higher' human capacities, and the failure to exercise those capacities is a deviation from human nature. In his seminal essay "Two Concepts of Liberty," Berlin explicitly associates Kant with this tradition that distinguishes between a "lower" or "base" human nature and a "higher", "real", or "true" human nature that is rooted in reason (178-81). According to Berlin, authoritarian political

regimes have justified themselves precisely through appeals to the idea that the lives of some members of their populations have deviated from or violate this “true,” rational nature. Ignoring whether or not Berlin’s claim about the history of political rhetoric is correct, it is true that Kant, along with most philosophers in the Western canon, distinguishes between ‘higher’ and ‘lower’ faculties, and refers to reason as a ‘higher’ faculty. It is a short step to the claim that that acting a way that reason requires counts as following our ‘higher’ nature. Moreover, Butler explicitly argues that that the idea of a ‘higher’ nature that one follows when acting virtuously and violates when acting viciously entails, or is perhaps even identical to, his account of the whole of human nature, writing that “the terms higher and better, imply a relation or respect of parts to each other; and these relative parts being in one and the same nature, form a constitution and are the very idea of it” (P: 16).²²

Put succinctly, the objection then runs as follows: (1) Ethical self-legislation is a ‘higher’ human capacity. (2) The proficient exercise of ethical self-legislation is virtue. (3) Because the exercise of ‘higher’ human capacities counts as following our ‘higher’ human nature, and (4) ‘higher’ human nature either entails or is identical to Butlerian human nature as a whole, (5) Kantian virtue consists in following human nature in Butler’s sense.

I contend that this objection fails because premise (4) is false. Although Butler often uses the term ‘true’ nature interchangeably with ‘whole’ nature, and the terms ‘true’ and ‘higher’

²² Darwall distinguishes between Butler’s conception of human nature as a *constitutional order*, which involves only the distinction between ‘higher’ and ‘lower’ parts of human nature, from his conception of human nature as a *teleological system* (1992: 223- 4). As Darwall notes, however, Butler treats them as interchangeable. For example, in introducing his watch analogy, which the primary passage in which he explains his conception of teleological systems, he asks us to state to ourselves “exactly the idea of a system, economy, or constitution of any particular nature” (P: 14). He continues throughout the analogy to use the terms ‘system’ and ‘constitution’ interchangeably, claiming that when we consider the parts of human nature individually, we do not arrive at “an idea of the system or constitution of this nature,” but that when we consider “the relations which the several appetites and passions in the inward frame have to each other...we get the idea of the system or constitution of human nature” (P: 14). The fact that Butler treats them as interchangeable gives us good reason to think that when he uses the term ‘constitution,’ we can read him as referring to the notion of teleological systems outlined above.

nature are often used interchangeably by writers such as Berlin, Butler's account of our 'true' or 'whole' nature is distinct from Kant's account of our 'higher' nature. That is, accounts of human nature that distinguish between its 'higher' and 'lower' parts, such as Kant's, and those that regard it as a teleological system are not identical. Darwall's distinction between a constitutional order and a teleological system is helpful here.²³ Butler uses the terms 'constitution' and 'system' interchangeably and, I contend, relies heavily on his systematic or teleological conception of human nature in his arguments for the authority of conscience and his Stoic conception of virtue. But Darwall rightly notes that the terms 'constitutional order' and 'teleological system' can be used to signify two distinct things. A constitutional order is characterized by a distinction between authoritative and subordinate parts. In a constitutional order, some part or parts of a whole have legitimate authority over other parts. The distinction between the authoritative and subordinate parts of a constitutional order is often expressed in the language of 'higher' and 'lower' parts. Kant has a conception of human nature that can be characterized metaphorically as a constitutional order. Reason has legitimate authority over other parts of human nature such as passion, inclination, and affect, and is a 'higher' part of human nature in this sense.

This conception of human nature is, however, distinct from a teleological conception of human nature such as Butler's. The crucial differences here pertain to Butler's claims that the whole of a teleological system must be disposed towards performing its proper function that it is and that the proper function of the parts must contribute to the function of the whole. Kantian human nature, I have argued, does not fulfill the latter condition—its parts are not appropriately

²³ Darwall claims that, although Butler uses 'constitution' and 'system' seemingly interchangeably when referring to human nature, clarifying Butler's argument for the authority of conscience requires distinguishing the two terms, and showing that the authority of conscience relies on the conception of human nature as a constitutional order (Darwall 1992: 233-4).

characterized in terms of their contribution to the function of the whole. I cannot defend the claim here, but I do not see any good reason to think that the whole of Kantian human nature has any function at all. It cannot be the case then that Butler's 'whole' or 'true' nature is identical to Kant's 'higher' nature. Kant does, however, distinguish between 'higher' and 'lower' parts of human nature and is warranted in doing so even though he does not share Butler's teleological conception of human nature. As we have seen, Butler's argument that the function of parts of human nature such as self-love and resentment is characterized by the authority of conscience did not follow directly from the premise that conscience is authoritative. This argument relied on the separate conviction that human nature is a system in a sense akin the sense in which a watch is a system. Not only is the idea of a 'higher' nature different from Butler's idea of human nature as a teleological system, but the former also does not, by itself, entail the latter.

The referent of 'human nature' in Butler's claim that virtue consists in following human nature is the account of human nature outlined above that he shares with the Stoics. Whatever sense in which it might be fair to say that Kantian virtue consists in following human nature, therefore, is not the sense relevant to Butler's claim that virtue consists in following human nature. Because Butler's teleological conception of human nature is essential to his Stoic conception of virtue, and Kant does not share Butler's teleological conception of human nature, Butler's stoic conception of virtue is unavailable to Kant. Butlerian virtue consists in following human nature, and it therefore it tends towards psychic harmony and happiness. Kantian virtue, on the other hand, is appropriately characterized in terms of a struggle, or perhaps even a battle, between reason and inclination, and offers little hope for psychic harmony or happiness.

8 CONCLUSION

I have argued that Kant and Butler's moral theories do not converge at the points we might expect them to. Both Kant and Butler think that autonomy, under some description, justifies claims about moral obligation. For Butler, this means that careful analysis of human nature shows that the authority of conscience characterizes the whole system of human nature, and that we are therefore a law to ourselves. Kant rejects each key element of this conception of autonomy. For Kant, there is no valid inference from the authority of conscience, or in his terminology 'reason,' to a description of the whole of human nature, and obligation cannot be justified by *any* facts that are particular to human nature. Kantian obligation is instead justified by the autonomy of reason, which is not dependent on its instantiation in human nature for the force of its claims. Finally, because Kant rejects Butler's broadly Stoic conception of human nature, Kantian virtue cannot be understood as following human nature in the sense relevant to Butler. That is, Butlerian Stoic virtue is unavailable to Kant.

I have avoided making any claims about whether I find Butler or Kant's arguments more convincing. Clearly, Kantian autonomy has enjoyed more historical prominence than Butlerian autonomy. One could venture any number of explanations for this fact—Kant's overwhelming influence in other areas of philosophy and Butler's homiletic form both come readily to mind—but I think we should avoid in attributing the relative prominence of Kantian autonomy to the superiority of Kant's arguments too quickly.

To begin with, they do not use terms in precisely the same way, and the relative success of their arguments hang, at least in part, on whose uses of the relevant terms we prefer. Both Kant and Butler think that persons are morally obligated when they have overriding moral reason to perform some action. But Kant's argument against Butlerian autonomy hangs on a claim about

the nature of obligation that would likely seem strange to Butler, namely, that the ground of obligation must be *absolutely* universal. Butler's goal in his *Sermons* is largely rhetorical—he wants to persuade the members of his church to behave virtuously. If defending a concept of obligation that is not absolutely universal, but is instead particular to human beings, is sufficient for achieving this goal, then Butler has no good reason to concern himself with Kant's objections. And if we are sympathetic to Butler's more localized conception of obligation, then we might fairly think that Kant's arguments against Butler miss the mark, that they defend a concept of obligation that is unnecessary for making sense of human moral life.

It might be tempting to think that Butlerian autonomy is not a living option for contemporary philosophers because of its reliance on a teleological conception of human nature. This thought is part of Darwall's reason for reconstructing Butler's argument for the authority of conscience in a way that avoids relying on teleological principles. However, as Wedgewood has pointed out, the familiar narrative according to which modern history has seen the victory of mechanism over teleology is not quite accurate. Both Butler and Kant are participants in a debate about the how far mechanistic explanations can extend, and this debate continues into the present. It has been argued in more recent years that even Darwin, who is often regarded as driving teleology out of its last refuge, namely biological theory, retained important elements of teleology in his theory of evolution.²⁴ And whether teleological principles ought to be extend in to our accounts of human practical psychology is still a debate worth having. A Butlerian account of human nature—and a conception of autonomy based on it—is then perhaps more plausible for a contemporary philosopher to ascribe to than it might initially seem.

²⁴ (Lenox 1993)

Regardless of whether we ultimately find Butlerian or Kantian autonomy more convincing, I hope to have motivated two claims. First, which conception of autonomy is more convincing is not an idle question—it has ramifications for what further positions are and are not available to us. Second, Kant’s is not the only concept of autonomy that is historically available to us. Contemporary philosophers who are concerned with defending a view moral life centered on practical agency’s capacity for self-legislation, but find Kantian ethics overly austere, might draw fruitful inspiration from Butler’s *Sermons*.

REFERENCES

- Butler, J. (2017) *Fifteen Sermons and Other Writings on Ethics*. Ed. McNaughton, D. Oxford: Oxford University Press.
- Brownsey, P.F. (1995) 'Butler's Argument for the Natural Authority of Conscience', *British Journal for the History of Philosophy* 3: (1): Pp. 57-87
- Darwall, S. (1991) 'Conscience as Self-Authorizing in Butler's Ethics' in *Joseph Butler's Moral and Religious Thought*. Ed. Cunliffe, C. New York: Oxford University Press.
- Frierson, P. (2013) *What is the Human Being?*. New York: Routledge.
- Hobbes, T. (1994) *Leviathan*. Ed. Curley, E. Indianapolis: Hackett Publishing Company.
- Hume, D. (1993) *Enquiry Concerning Human Understanding*. Ed. Steinberg, E. Indianapolis: Hackett Publishing Company.
- Kant, I. (1922) *Gesammelte Schriften*. Ed. Königlich Preussischen Akademie der Wissenschaften. Berlin and Leipzig: de Gruyter.
- Kant, I. (1996a) *Groundwork for the Metaphysics of Morals*. Transl. Mary Gregor. New York: Cambridge University Press.
- Kant, I. (1996b) *Metaphysics of Morals*. Transl. Mary Gregor. New York: Cambridge University Press.
- Kant, I. (1996c) *Critique of Practical Reason*. Transl. Mary Gregor. New York: Cambridge University Press.
- Kant, I. (1998) *Critique of Pure Reason*. Transl. Paul Guyer and Allen W. Wood. New York: Cambridge University Press.
- Lenox, J. (1993) 'Darwin was a Teleologist', *Philosophy and Biology*. 8 Pp. 409-21.

Long, A. (2003) 'Stoicism in the Philosophical Tradition: Spinoza, Lipsius, Butler' in *Cambridge Companion to the Stoics*. Ed. Inwood, B. New York: Cambridge University Press.

Mandeville, B. (1997) *The Fable of the Bees and Other Writings*. Ed. Hundert, E.J. Indianapolis: Hackett Publishing Company.

Penelhum, T. (1985) *Butler*. London: Routledge & Kegan Paul plc.

Reich, K. (1939) 'Kant and Greek Ethics', *Mind*. 48(191) Pp. 338-354.

Sturgeon, N. (1976) 'Nature and Conscience in Butler's Ethics', *Philosophical Review*. 85(3) Pp. 316-356.

Wedgewood, R. (2007) 'Butler on Virtue, Self Interest, and Human Nature' in *Morality and Self-Interest* Ed. Bloomfield, P. Oxford: Oxford University Press.

Wilson, E. (2015) 'Self-Legislation and Self-Command in Kant's Ethics', *Pacific Philosophical Quarterly*. 96(2) Pp. 256-278.