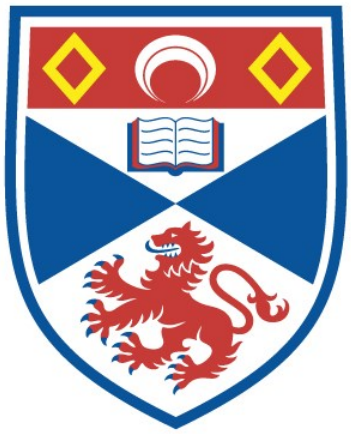
Kantian constitutivism and the limits of agency

Michael Paul Frank

A thesis submitted for the degree of MPhil
at the
University of St Andrews



2025

Candidate's declaration

I, Michael Paul Frank, do hereby certify that this thesis, submitted for the degree of MPhil, which is approximately 40,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2022.

I confirm that no funding was received for this work.

Date 26/09/2024 Signature of candidate: Michael Frank

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of MPhil in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

26 IX 2024

Date Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to

migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Michael Paul Frank, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 26/09/2024 Signature of candidate: Michael Frank

26 IX 2024

Date Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Michael Paul Frank, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date 26/09/2024 Signature of candidate: Michael Frank

General acknowledgements

I would like to thank my supervisors, Professor Jens Timmermann and Dr. Justin Snedegar, for reading countless drafts of this current work. I would also like to acknowledge Professor Cecil L. Eubanks, who introduced me to the works of Kant and Rawls while I was still an undergraduate student. Any mistakes in this work are my own.

Abstract:

The philosophical method referred to as constitutivism regards some norms as constitutive of the active exercise of a particular capacity. The distinctive feature of Korsgaard's Self-Constitution Constitutivism (SCC) is the way that it explains and justifies normativity. The SCC theory treats the principle of practical reason as constitutive of the capacity for agency.

While I take the SCC theory to be an ingenious attempt at providing an explanatory and justificatory ground for moral and non-moral normativity, the account it provides has come under criticism both from inside and outside the Kantian tradition. Those inside the Kantian tradition often criticize the view for its tendentious interpretation of practical reason. Those outside the Kantian tradition often criticize the view for offering an account of normativity that is insufficient to ground anything like categorical moral authority.

As a Kantian, with broadly sympathetic constructivist leanings, I aim to answer the criticisms of those from outside the Kantian tradition by taking on board the criticisms of Kantians. Specifically, I address three core problems and themes. First, I look to the problem of motivational ambivalence. Second, I discuss the relational theory of value. And, third, I discuss the objection that constitutivism provides insufficient ground for categorical moral authority.

I conclude by making a few suggestions that allow philosophers of different types to move past these debates. First, for philosophers who align with the SCC theory, I suggest that, instead of speaking of a unified capacity for rational agency, the theory would be better served by speaking of several capacities, each with their own rationally structured function. Second, for Kantians more broadly, I show what role constitutivism is playing in Kant's practical philosophy. From this, it becomes clear that constitutivism cannot answer all metaethical questions.

Table of Contents

Chapter V

Chapter One

1. The Constructivist Method

Constructivism, in ethics and metaethics, is a method whereby objective conclusions can be reached through an acceptable procedure or some normative standpoint. Types of constructivism largely vary depending upon what the theory is meant to model, explain, or justify. The classical version was proposed by John Rawls and attributed to Immanuel Kant. The Rawlsian version begins with certain irreducibly normative elements, which are then combined with a procedure of construction for the purpose of reaching an objective conclusion about the target of inquiry.

In the classic interpretation, the irreducibly normative elements at play are the procedure itself, a conception of personhood, and requirements of practical reason. Once the elements are combined, the constructivist's answer is meant to be objective but response dependent. The answer is objective because it is assumed that all would accept the conclusion reached from within the confines of the procedure. The answer is response dependent because the objective conclusion is not independent of what the agent does choose. Kant, it was argued, should be classified as a constructivist because he took seriously the idea that ethical statements can be true and yet dependent on the judgement of the person.[1]

Constructivist philosophers like Sharon Street (2010) suggest that the truly distinct aspect of constructivism is not the procedure of construction, but the normative standpoint from within which agents reason about what to do. One either reasons oneself into the normative standpoint, in which case the standpoint itself is constructed from minimally normative elements, or one takes certain rules to be constitutive of moral experience, such that every moral agent aims at following these rules. The former option takes construction to go all the way down. The latter option is termed constitutivism, which I will describe further in the next section.

2. The Constitutivist Solution

[1] See (Rawls, 1980).

The constitutivist justification of normativity typically starts from the assumption that certain rules constitute the activity of moral reasoning. These rules provide a function to whatever activity it is meant to constitute. For example, if I were to claim that the proscription of lying is constitutive of human moral cognition, then I would assume that this rule constitutes the activity of human moral reasoning. The constitutive function is meant to operate on two levels: the descriptive level and the normative level. On the level of description, the function dictates how the activity in ideal conditions necessarily will proceed. On the normative level, the function dictates how the activity governed by the rule should be performed.

There are several types of constitutivism that dominate current debates. One type is naturalist constitutivism, whereby a normative function is ascribable to some class of beings that belong to a natural kind. Another type is rationalist constitutivism, whereby a normative function is ascribable to some class of beings by virtue of these beings' rational powers.[2] Despite the differences, both versions of the constitutivist view emphasize what is sometimes termed the philosophy of capacities.

Rationalist constitutivism often stresses certain rational capacities. On this view, rationality structures the capacity to perform some activity by representing the capacity's purpose or ideal function. Suppose, for example, there is a rational being with the constitutive capacity to follow the rules of chess. By virtue of rationality representing the capacity's functional rules, the chess player must aim at playing chess well. Suppose that rationality, in this instance, represents the purpose of chess playing to be checkmate. In that case, the rational being aims at checkmate because this is the capacity's representative function.

Much recent philosophical work in constructivism has focused on the theory's Kantian pedigree. Kantian constructivists have been focusing on Kant's philosophical justification of normativity. Often, these constructivists ask what justification might be given for the normative elements of the original constructivist theory. Many of these

[2] See (Fix, 2021, 2023) on the division of constitutivism into naturalist and non-naturalist types. There is a difficult question here about whether rationalism can be squared with naturalism. See (Schafer, 2023, pp.54-55) for some thoughts about how Kant's view of capacities might be accommodated by a "naturalist" world view.

philosophers attribute a form of rationalist constitutivism to Kant as an answer to the question of normative justification.

Prominent constitutivist interpretations of Kant typically fall into two main categories. Some Kantians, such as Stephen Engstrom (2009), focus on the rational activity of judgement. Often, these philosopher model practical judgements on Kant's notion of theoretical judgements of the understanding. These philosophers might be termed intellectualists. In contrast, other philosophers interpret practical judgement as dependent upon the activity of rational agency. This group interprets Kant's comments about the rational will as an emphasis on rational choice and action. Since rational agency is the focus of this second grouping, they might be termed pragmatists.

Ultimately, I say very little about the intellectualists in what follows. Instead, I focus on one Kantian interpretation with largely pragmatist leanings. Specifically, I focus on Christine Korsgaard's (1996a, 1996b, 2008, 2009) interpretation of Kant, which can be termed Self-Constitution Constitutivism (SCC). I take the distinctive feature of Korsgaard's view to be the way that she transposes the original constructivist elements to the standpoint of rational deliberation. In essence, Korsgaard takes these normative elements to constitute the identity of agency, which is an identity that it is assumed all share. On this view, rationality represents the function of rational agency, such that agents must aim at constituting themselves, including their actions and mental states, as fully rational or virtuous.

3. Korsgaard's Constitutivism

Before I engage in my argument, I will motivate my decision to focus on Korsgaard's version of Kantian constitutivism. First, I focus on this version because it has been, without a doubt, one of the most influential Kant interpretations within late-twentieth and early-twenty-first century anglophone philosophy. Second, the pragmatic emphasis on agency has made the SCC view seem like a more plausible justification of normativity than Kantian transcendental idealism. Korsgaard's answer to what she terms "the normative problem" (1996b, pp.92–94)—a problem which questions how it is that rational beings have authoritative reason for action—is a first-personal emphasis on identity. We have reason to act because, in the first-person perspective, we must act,

and these reasons hold normative authority because we rationally impose these reasons on ourselves.

Assuredly, Korsgaard's answer to the normative problem is very Kantian in character. However, one main criticism of this framing is that it cannot purchase the desired notion of normative authority. There are several reasons to question Korsgaard's conclusion that a justification of categorical moral authority—without which the theory would not be truly Kantian—will be forthcoming from the first-personal identity of rational agency. My aim is to explore what these reasons are and to ask whether they hold argumentative force.

In order to facilitate the investigation at hand, I look to several prominent criticisms of the SCC view. Notably, Korsgaard has been criticized both from within the Kantian tradition and from without, by philosophers who hold different assumptions. I look to both types of criticisms. My hope is that we can accommodate the points made by other Kantians in order to answer the criticism made by non-Kantians.

To be sure, I believe there is much that Korsgaard's view gets right about Kant. However, my stance in what follows will largely be critical. My aim with regard to Korsgaard's work is twofold: I hope to present Korsgaard's view charitably. And, I hope to provide criticism which can help those who subscribe to the SCC theory move forward.

Before I continue, I should add a quick note about my engagement with Korsgaard's and Kant's works. Both of these philosophers have been incredibly prolific, and as such, I inevitably focus on certain texts more than others. For Kant, I largely focus on the period of his writing that takes place roughly from 1781-1788. This encompasses the start of Kant's critical period, with publication of the *Critique of Pure Reason*, and the development of Kant's mature ethical thought with the publications of the *Groundwork of the Metaphysics of Morals* and the *Critique of Practical Reason*.[3] I

[3] Henceforth, I refer to the *Critique of Pure Reason* as the *First Critique*, the *Groundwork of the Metaphysics of Morals* as the *Groundwork*, and the *Critique of Practical Reason* as the *Second Critique*. All citations of Kant's work follow the standard practice of providing the Akademie edition volume followed by the Akademie page numbers. Also standardly, I follow the practice of citing the *First Critique* by referencing the A-edition page number followed by the B-edition page number.

engage with Kant's later works as well, but most debates about Kantian constitutivism focus on this period of Kant's writing in particular.

My engagement with Korsgaard's ideas will discuss all periods of her work thus far. Notably, though, I do not engage with Korsgaard's work on animal ethics.[4] However, I do engage with some of Korsgaard's other work from that point in her career. I engage mostly with the period of Korsgaard's writings that span from *Creating the Kingdom of Ends* (1996a) to *Self-Constitution: Agency, Identity, and Integrity* (2009) because these are the works which have had the most impact on the constitutivist debates. Sometimes I refer to "the classical" or "orthodox interpretation of Kant" as a view to be contrasted with Korsgaard's SCC theory. I do not intend this name to be a value judgement. Rather, I use it as a scholarly category for the purpose of argument. Korsgaard (1995) has provided a defense of her distinctive methodology for interpreting historical philosophers elsewhere. My use of the term "orthodox" is only meant to signal this difference in methodology, although at times the difference in methodology will also signal an important philosophical divergence.

4. The Argument

My main argument will be that the function of rational agency is insufficient to provide a ground for categorical moral authority. In contrast, I will set forth a different interpretation of Kant which does not rely on the identity of rational agency as a philosophical justification. Rather, my interpretation of Kant suggests that the principle of autonomy, or pure reason in its practical use, constitutes the capacity to make practical judgements that determine the moral law. In contrast to Korsgaard, I do not take practical judgements to serve, foremost, the function of constituting the rational agent. In effect, my interpretation emphasizes less the first-personal question of efficiency, and I emphasize, instead, the capacity to make moral judgements.

Largely, Korsgaard's framing has been met with skepticism because it takes two distinct projects to be conjoined. The SCC theory takes the projects of normative explanation and the justification of moral authority to be the same. The answer to both

[4] While this debate is important and assuredly does have implications for the current debate about constitutivism, I do not have a considered view about how to extend Kantian obligations to animals.

of these questions are said to come from the universal identity of rational agency. This has, in turn, allowed for skeptical arguments which question the viability of this project. In contrast, I do not interpret Kant as attempting to answer these kinds of skeptics. Kant, on my view, takes these projects of normative explanation and moral justification to be distinct.

Before I turn to that worry, though, I address problems that have been raised for the SCC theory's notion of moral motivation and moral value. In particular, the SCC theory has been criticized for its specific notion of moral rationalism. In Chapter Two, I argue that a common reading of Korsgaard's rationalism is mistaken, but I also argue that the SCC theory is subject to a particular type of motivational bootstrapping worry.

Chapter Three deals with the relational theory of value that Korsgaard has attributed to Kant. Most of Chapter Three focuses on exegetical questions. The chapter allows for a clarified presentation of the SCC theory's main assumptions. In effect, Chapter Three acts as a segue to the main criticisms at issue in Chapter Four, where I deal with the problem of insufficient moral authority. I conclude by making certain prescriptions which might help those who subscribe to the SCC theory to move past the counterarguments that I raise.

Chapter Two

1. Introduction

Philosophers sometimes assume that an agent must perform a moral act wholeheartedly because only a wholehearted action is one that is fully authored by the agent. If an agent acts morally but ambivalently, then the agent does not fully possess the action as his own. The act is done in accord with duty, but not *for the sake of* one's duty. In the language of practical reasoning, the act is not done for the right reason. Korsgaard's SCC theory, as an interpretation of Kant's practical philosophy, follows Kant on this point, though this is left as an implicit assumption of the theory.

Some philosophers, in contrast to those who support wholehearted action, have suggested that there are certain cases where ambivalence is the only justifiable response. This argument questions the extensional adequacy of wholehearted action theories. If the views which support wholeheartedness cannot account for justifiable ambivalence cases, then the theory needs modification in order to capture this facet of moral psychology.

I will suggest that Korsgaard's SCC theory cannot account for cases of justifiable ambivalence. Furthermore, I will suggest that this explanatory failure is largely due to the specific form of moral rationalism ascribable to Korsgaard. In section 2.1, I outline the views on wholeheartedness and ambivalence. In section 2.2, I discuss Korsgaard's interpretation of maxim endorsement. In sections 3.1 and 3.2, I will deny one interpretation of Korsgaard's rationalism, but I suggest that the better interpretation still leads to a bootstrapping criticism. Ultimately, I argue that Korsgaard cannot accommodate cases of justifiable ambivalence.

2.1 The Native Hue of Resolution Sicklied over with Thought

The phenomenon of *ambivalence* has been of recent philosophical interest in moral psychology (Brunero, 2021; Gunnarsson, 2013). Self-Constitution theorists have taken up this discussion as one of central import for their target discourse, the question of how the agent or person constitutes a coherent self when faced with disparate, often contradictory, courses of action. Harry Frankfurt (1988), for example, argues that one

must be *wholehearted* in one's decisions and actions, and to be anything less than wholehearted is to fail in constituting oneself. Wholeheartedness is a necessary condition of unified agency, or so the reasoning goes, because to be halfhearted, for example, is to divide oneself and be subject to the vicissitudes of contingent influence. Asserting one's agency, under this line of thought, requires that one take control—or authorship, insofar as one is capable—of one's own actions: to be halfhearted about some chosen course of action is to surrender one's capacity for choice to something other than self-control.

Ambivalence in one's actions is antithetical to the requirement of wholeheartedness. As such, ambivalence breaks a norm which the Self-Constitution Theorist describes as governing agency. Some philosophers focus on a general norm of wholeheartedness. Similarly, Korsgaard's SCC theory assumes that prudential normativity is entailed by the principle of practical reason.[5] It follows that if an agent is not wholeheartedly acting, then the agent is contravening his own principle. In this section, I will focus my attention on cases of practical normativity in general. Later, I will focus my attention on cases of morally valuable actions. Since the SCC theory assumes that prudential reasons and moral reasons share the same source, the conclusions I reach in this section will generalize to cases of moral reasoning.

To illustrate why some philosophers believe agency requires a norm like wholeheartedness, one can look to cases in which some agent must choose between distinct life paths. For example, Jones is offered an academic position at a top university. Jones is also offered a high-paying job at a tech firm. The academic job will offer more prestige in a field that Jones has spent much of his life studying. On top of this, Jones truly enjoys the life of a researcher. In contrast, the tech firm offers a higher salary, a better quality of life, and a greater amount of leisure time. Jones cannot accept both jobs. Failing to choose either would constitute the overall worst outcome.

Assume now that Jones chooses ambivalently to work in the tech firm. Jones chooses the tech firm, but he wishes he had chosen differently. At the very least, Jones

[5] Some philosophers (Korsgaard, 1996a, pp.242–243; O'Neill, 1989, pp.28–50) interpret the "primacy of practical reason" as a view suggesting that the function of practical reason is the source of theoretical reason. See (Schafer, 2023) for an interpretation which balances the practical and theoretical functions of reason without obscuring the distinction.

is less than certain that his choice was the correct one. In this case, Jones faces a scenario in which his practical reasoning is pulled toward an act that is no longer open: he has turned down the academic job, and the job has been offered to another candidate. The chosen job, that of the tech firm, is a life path with which Jones cannot fully identify. If ambivalence about one's choices makes an agent feel like less than an author of his own choices, then ambivalence contradicts what the Self-Constitution theorist takes to be fundamental to agency.

The above illustrates why Self-Constitution theorists make wholeheartedness a norm of agency. But critics of the wholeheartedness view often claim that ambivalence is not antithetical to the norms of agency (Brunero, 2021; Gunnarsson, 2013). In contrast to the Self-Constitution theorists, some philosophers claim that ambivalence can be a justifiable expression of agency and that certain cases call for an ambivalent response. In these cases, anything other than ambivalence would fail to express the agent's authorship of himself or a truthful relationship to his own emotions. These philosophers readily point to these justifiable ambivalence cases to show that wholeheartedness is an artificial constraint imposed from outside the plausible course of an agent's own reasoning. Thus, these philosophers conclude, wholeheartedness cannot be a norm internal to or constitutive of the activity of practical reasoning.

Before proceeding further, it will be useful to distinguish between two common conceptions of ambivalence. This distinction can be made clear by turning to the entry on ambivalence in "general contexts" in the *Oxford English Dictionary*. The entry defines ambivalence as, "[1.] The condition of having contradictory or mixed feelings, attitudes, or urges regarding a person or thing. Also: [2.] the condition of being undecided about a viewpoint or course of action".[6] I take the first definition to correspond with *mental ambivalence*, or conflict between attitudes which point toward the same state of affairs. The second definition, on the other hand, corresponds with *motivational ambivalence*, or indecision in an agent's actions. These are interrelated concepts; hence, the same word refers to both. However, I take these notions to be importantly distinct.

[6] Oxford English Dictionary, s.v. "ambivalence (n.), sense 1.b," September 2024, https://doi.org/10.1093/OED/3055940960.

For example, an agent could express mental ambivalence without expressing motivational ambivalence. It might be the case that Jones feels conflicted about his boss. Jones regards his boss as both a savvy businessman and an unscrupulous tyrant. In fact, Jones expresses his conflicting mental states through critical conversation with his friends. In no way is Jones motivationally undecided in this case. And yet, counterfactually, if Jones were to express motivational indecision about working for his boss—by, for example, contemplating whether to leave for work in the morning or whether to quit—then this presumably points to a case of mental ambivalence as well.[7] It is only because Jones *feels ambivalent* that he *acts with indecision*. Thus, there can be scenarios in which an agent expresses mental ambivalence without motivational ambivalence, but the converse is not true.

It is at least plausible that mental ambivalence should be temporally prior to motivational ambivalence when diachronically modeling the activity of practical reasoning. I will focus on motivational ambivalence for two reasons. First, I am primarily concerned with action. And secondly, I take motivational ambivalence cases to cover cases of mental ambivalence as well.

With the types of ambivalence distinguished, I will now provide a case of justifiable ambivalence. Philosophers who argue in favor of justifiable ambivalence might motivate the argument by looking to the example of Prince Hamlet, a classic literary case of ambivalence. Called back to the royal court of Elsinore from his studies at Wittenberg, Hamlet is met with a political and personal crisis that fractures his psyche. The general structure of Hamlet's case is one in which an agent has values that he wants to pursue, but the pursuit of one value is certain to undermine the possibility of pursuing the other.[8]

Take, as an example, Act Three, Scene Three, wherein Hamlet famously has a chance to revenge himself upon King Claudius for the murder of his father (pp.167–169).[9] Hamlet chances upon Claudius alone, though fate would have it that Claudius is,

---

[7] When discussing *practical rationality*, Kantian theorists often subscribe to a strong notion of internal authority, with some important caveats concerning the limits of *theoretical self-knowledge*. My distinction relies on, at least, a broadly sympathetic framing of the epistemology of practical self-knowledge.
[8] The argument resembles the literature on moral dilemmas. See (Timmermann, 2013) for an overview. I follow Timmermann in suggesting that Kant is able to accommodate cases of agent regret.
[9] From *The Folger Shakespeare*.

in that moment, a supplicant, praying for forgiveness. Or, at least, this is how the scene appears to Hamlet, who must decide whether to kill his enemy during prayer or wait until his enemy is engaged in a vicious act. On the one hand, Hamlet desires revenge. On the other hand, Hamlet is afraid of the uncertain future which revenge might bring about. This is the case of mental ambivalence at play in his deliberations. [10]

While the case of mental ambivalence is important for understanding the background conditions which shape the specific case of deliberative conflict at issue, it is important to look specifically at the motivational ambivalence which emerges in this scene. This second notion of ambivalence is clarified by the nature of Hamlet's *rationalizations* about the possible effects of his action. Instead of acting with determination, Hamlet provides himself with an excuse: to kill Claudius during prayer would be to ensure that Claudius dies a virtuous man. But, is Hamlet's unwillingness to take revenge a sincere moment of faith? Only two scenes prior, in Act Three, Scene One, Hamlet expresses his uncertainty in the existence of God and heaven (pp.127–129).[11] A more plausible reading would suggest that Hamlet, in failing to take decisive action, is expressing his conflicting attitudes about the prospect of killing Claudius, not suddenly voicing a sincere belief in religious doctrine. Ultimately, Hamlet sheaths his sword and vows to kill Claudius at another time.

Although Hamlet fails to act in this scene, it is precisely the failure to act which is an authentic expression of Hamlet's fraught psychology. Wholehearted determination in favor of killing Claudius would contradict Hamlet's own justified fears. Surely, if Hamlet were to act decisively, the kingdom of Elsinore would be thrown into turmoil and Hamlet himself would face danger. But, equally, wholehearted determination in favor of allowing Claudius to go free would contradict Hamlet's justified desire for revenge.[12] In this scenario, there is no case where wholehearted action can be an authentic expression of the agent.

---

[10] Famously, *Hamlet* provokes myriad interpretations. I am bracketing alternative interpretations.

[11] Hamlet refers to death as "the undiscovered country from whose bourn no traveler returns" and says it "puzzles the will" (lines 87-88).

[12] Note again that I am here talking about a non-moral case of instrumental reasoning. When I use the term "justification," in this case, it simply refers to something like *instrumentally rational*, where this means that instrumental reason can be given in favor of the action.

Now, if we remove the self-imposed limitation on discussions of moral ambivalence, one might argue that the Hamlet case cannot really serve as a criticism against the SCC theory because revenge is immoral. The argument would suggest that wholeheartedness is really only *justified* in cases where the agent has an all things considered best reason to perform the action in question. Moral reasons defeat non-moral reasons when they conflict. Reasons for revenge are non-moral reasons. Therefore, an agent cannot act wholeheartedly in favor of revenge.

While this argument is sound, it suggests a model of practical reasoning whereby a reason for decisive action can also bootstrap the agent out of mental ambivalence. This is essentially because the SCC theory relies on an account of reasoning whereby mental states serve as one of the components that agents use to construct reasons for action. If this is the case, then it does not seem possible to distinguish between mental ambivalence and motivational ambivalence.

For example, if we apply the SCC account of reasoning to the Hamlet case, then Hamlet has a *reason* for action when he takes up certain salient features of the world and constructs them into a reason. Hamlet's mental states are part of the constructive procedure. Therefore, if there is a scenario in which a moral reason defeats a non-moral reason, then two conclusions must be true from these assumptions. First, the moral reason is a reason to perform the *moral action*. Second, since the agent's mental states are part of the reason's content, then the moral reason must also serve as a reason to direct *fitting attitudes*, such as moral approbation, toward the action. Later in the chapter, I will explain these assumptions further, but for now, I assume that this information is sufficient to make my point.

If what I have just argued is correct, then the argument against the validity of justifiable ambivalence cases cannot work. Moral reason, under this model, entails that an agent is irrational if he does the moral action and simultaneously holds attitudes of regret or wishes that circumstances could have been different. With these assumptions, it would be irrational for Hamlet to be motivationally wholehearted and also feel regret. In other words, motivational ambivalence and mental ambivalence are not separate concepts for the SCC theory.

Is it completely irrational to hold a regretful attitude toward a state of affairs in which you are incapable of satisfying one of your desires? It is unclear that bootstrapping oneself out of certain attitudes is really the remit of rationality. So long as the agent is not motivationally ambivalent, then it is not irrational for the agent to express mental ambivalence. But, the SCC theory can only tell Hamlet that he should undermine his own attitudes. In contrast, we need a model of moral motivation that allows us to say that Hamlet *should* act from duty, but that this action might not necessarily change Hamlet's ambivalent attitudes. And even if Hamlet does not change his ambivalent attitudes, he is still *moral and rational* so long as he acts for the right reason. In other words, to take this criticism on board, we need a theory that accommodates moral motivation but also allows for the expression of ambivalent attitudes.

2.2 The Form of Action and Self-Conscious Endorsement

To understand why it is that Korsgaard commits herself to this specific version of moral motivation, we must first understand how Korsgaard interprets the Kantian idea of a maxim and its endorsement. Foremost, for Korgaard, a maxim is the form that act types must take (2008, pp.218; 2009, pp.15–16). Any token act must conform to some maxim, which describes the act type. The maxim is functional, in the sense that it is governed by a representative purpose or activity, and it is meant to be intelligible to other agents as a reason. If the maxim fails to qualify as a universalizable reason, then the maxim cannot be the best reason that one has.

An agent willing a maxim is an active process. There are distinct elements that the agent takes up and incorporates into his maxim. This is a version of the incorporation thesis at work in Kantian ethics. Typically, this thesis means that an agent must incorporate an incentive into his maxim. The SCC theory interprets this as incorporating the constituent components of an act into a maxim so that the maxim itself might be tested against a procedure of deliberation and then willed as a reason. To clarify how this works in practice, I will explain how it is that an agent constructs his reasons from the various constituent elements.

Every act has a purpose, or *telos*. For some acts—those which are chiefly concerned with the consequences brought about—the purpose is the effect, the causal change affected in the world. For other acts—those which are chiefly concerned with the rational validity of the action's form—the purpose is the action token, the activity of bringing about a specific instance of an act type which reason prescribes. Of course, this is not to say that agents concerned mainly with the rational validity of their own acts are not also concerned with certain consequences. This is just to say that certain actions are valued for their supposed universal form, rather than the consequences produced. The important point is essentially that a rational being can attend to the form of an action as well as the consequence produced.

The purpose of an action suggests itself to an agent through a certain psychological cause that Kant describes as a *Triebfeder*, literally a spring that triggers and sustains movement in a mechanistic system. Under classic interpretations of Kant, this is an element in the agent's psychology which plays the causal role. As such, the term *Triebfeder* is often translated as incentive. In the classic interpretation, incentives of inclination are purely psychological, non-cognitive representations which suggest possible candidates for pleasurable experience.

An agent might act on his pangs of hunger *because* he perceives a full Scottish breakfast, replete with fried eggs and potato scones, which suggests the incentive of pleasurable satiation. The classic interpretation takes the important point of emphasis to be the *pleasure* suggested by the incentive. By contrast, the SCC theory does not interpret incentives of inclination as merely psychological phenomena. Instead, this interpretation argues that incentives of inclination are *features of the world* which suggest themselves to the agent as the possible content for reasons for action. Specifically, Korsgaard describes incentives as "features of the objects of those inclinations that make them seem attractive and eligible" (2008, pp.109).[13] Later, I will suggest that this interpretation ultimately leads to a version of the bootstrapping worry because it cannot adequately differentiate between types of reasons. Specifically, it treats all practical reasons as serving the same function, where this function is to determine agency rationally.

[13] For a complementary interpretation, see also (Sussman, 2003).

Before I get to this criticism, I will continue my current analysis of maxim endorsement and motivation. If we focus on the classic interpretation of Kant's incentives of inclination, there is nothing thus far in the theory of motivation that might seem untoward to a philosopher who subscribes to the empiricist notion of psychology. Interpreted as a material causal force, incentives of inclination are like the Humean theory of motivation. The SCC theory, by contrast, takes this Humean psychological assumption to be an unlikely candidate for Kantian theory (Korsgaard, 1998; Sussman, 2003). Where Kant definitely diverges from the empiricist view is his insistence that an incentive might be formally causal, rather than material causal. This is Kant's famous dictum that pure reason can be practical of itself. For Kant, the form of a maxim suggests a special type of incentive: *respect* for the moral law. Respect is both a positive and negative incentive. It is negative in its ability to defeat the weaker incentives suggested by inclination. It is positive in its ability to provide the moral person, the person who acts from a sense of respect for the law, with a sense of moral worth.

While the SCC theory also argues that the form of a maxim is sufficient to motivate, I will suggest that Korsgaard's SCC theory has a very specific, and controversial, understanding of the way a maxim's material content interacts with its formal conditions. To see this, one can focus on the *double aspect theory of motivation*, a view which Korsgaard ascribes to Kant (2008, pp.174–206). If, as the SCC theory suggests, incentives of inclination are *mere candidates* for reasons, then alone these incentives are insufficient to motivate. The incentive must first be incorporated into a maxim, or constructed into a reason, for it to motivate. The double aspect theory of motivation suggests, in accord with its name, that motivation has two key elements: 1.) the incentive and 2.) the principle of volition. As we have already seen, the incentive suggests a course of action. The principle of volition, on the other hand, suggests the form that the incentive can take, such that it might be willed as a reason.

Under the SCC theory, the content of a reason comes from the incentives of inclination. Suppose an agent perceives a sugary soda. The fact that the agent perceives certain feature of this soda as thirst-quenching means that those features present themselves as *possible* reasons to drink the soda. But, as of yet, these features

are not reasons. Since the agent might choose to act on the appealing features of the soda or not, there is some aspect of the self over and above the reception of incentives. This is the second half of the dual aspect theory, the principle of choice or volition. This is where Korsgaard locates the moral worth of a maxim; to be a moral maxim, the maxim's incentive must be chosen through the correct deliberative procedure, otherwise the maxim cannot act as a reason and has no moral worth. This is also where the primary disagreement about the dual aspect theory lies. Chief among the concerns about this interpretation is the worry that the view conflates immoral action with mere unthinkingness. If I fail to reflect on whether the incentives in my maxim might be universalized, then I am guilty of a lapse in thought. I will have allowed my maxim to be entirely governed by the incentive of inclination.

The default principle of deliberation, then, is a principle of self-love, whereby maxims are formed from inclination alone. Self-love takes effect when the agent simply fails to reflect sufficiently. For example, a man is drinking at the pub, and he is angered because the football team he supports has lost. A rambunctious fan of the opposing team gloats gracelessly. The first man has a strong desire to throw a punch at rambunctious fan, and let us assume that he does throw said punch. The man has failed to reflect on whether his incentives are worthy of a universal maxim. Therefore, the man acts on a principle of self-love. Although alcohol in this instance might explain the man's deliberative failings, it cannot be an exculpatory factor.

While the drunken man example illustrates the SCC theory's notion of unreflective action, the moral person is the one who steps back from his incentives to see which, if any, might be put forth in a universalizable maxim. On this view, the categorical imperative is essentially a decision procedure, whereby incentives are tested for formal validity. An incentive can only serve as a sufficient reason for action if it passes the categorical imperative test. An agent has an inclination, wants to act on it, but first he must decide whether it can be put forth as a reason. First-order incentives put the agent on a certain track. But, the agent must choose: should the agent take his first-order incentive as sufficient, or should he prioritize the second-order incentive of *respect* for the formal validity of a maxim? Although there is much that is recognizably Kantian in this picture, I will suggest that it is a fundamentally flawed account. Respect

is not a higher-order incentive that tells the agent to submit his reasons to a test of sufficiency. Neither should the material content of a maxim be modeled entirely on desires of inclination. I will suggest that the SCC theory goes too far in homogenizing reasons and their sources when it models maxim formation on a procedure of deliberation.

The specific way that Korsgaard interprets Kantian rationalism leads to the problem of homogenous reasons suggested in the above paragraph. But, before I turn to the disagreements surrounding Kantian rationalism in the next section, I will address one further aspect of the psychology of maxim formation. Notably, Kantian ethics is about the validity of maxims. It is about the normative validity of one's own subjective description of an action. The maxim is distinct from the action itself, such that a person can be considered moral even if he intends something but fails in his execution. However, this does not mean that one's maxims are wholly divided from the capacity to act.[14] It is a core part of the Kantian view that maxims are descriptions that are judgeable candidates *for* action.

While I will bracket the question, for now, of whether or not Kant had a considered theory of *agency*, I believe Korsgaard's SCC theory gets at an important aspect of Kantian ethics. This is the idea that maxims are *to be enacted*. Rational beings are not wholly disinterested in the efficacy of their maxims, even though efficacy is not the primary metric by which maxims are judged to be valid. This is what I will call the *practical efficacy* view of maxims.[15] In willing a maxim, a rational being is self-consciously putting forth a specified description of the relevant act's features, and foremost among these features is the judgement that the act is, *at least*, taken to be efficacious and, *ideally*, morally good. The SCC theory gets this point right, but ultimately, in contrast to Korsgaard, I suggest that the *practical efficacy* view can be separated into three distinct requirements, which I take to cover Kant's notions of negative freedom, positive freedom, and merely instrumental normativity.

In contrast to my proposed three requirements, Korsgaard takes *practical efficacy* to be one theoretical package. As an implication of this, Korsgaard assumes

[14] The SCC theory would suggest that these are not distinct capacities at all.
[15] For Korsgaard's presentation of efficacy and autonomy, see (2009, pp.82–83).

that there is no such thing as a *mere* hypothetical imperative, and in addition, she takes instrumental rationality to be included under the heading of Kantian practical reason (2009, pp.70–72).[16] Korsgaard's argument in favor of this point is her "argument against particularistic willing," whereby she claims that maxims must conform to rational restrictions that are imposed by the agent by virtue of the agent's freedom (2008, pp.124). I will return to this point later, but for now suffice it to say that Korsgaard takes a will without restrictions to be a sort of simulacrum. A particularistic will is not really a will at all, but a collection of unguided inclinations. What the view gets correct is that there is *one* principle of practical reason, and hypothetical imperatives *should* be constrained by this principle if the agent is to execute permissible actions.[17] What I disagree with is the presentation of instrumental rationality as a facet of what Kant means by practical reason.[18] I will discuss this difference more later, but for now, I will suggest that there are purely pragmatic reasons to accept my tripartite distinction. Foremost, I will argue that conflating these requirements leads to a common misinterpretation about the types of moral rationalism which can be plausibly attributed to Korsgaard. Furthermore, if one likes, the distinction can be taken as a mere methodological assumption, and in that case, one can still subscribe to the SCC view while agreeing with my current argument. With this in mind, I turn to the three requirements below.

Requirement One corresponds to the notion of negative freedom, or the freedom one has in virtue of being undetermined by external influence. This requirement partially specifies the relevant meaning of efficacy in the *practical efficacy* view. The requirement might be stated like this: for some specified description of an action to qualify as a maxim available to choice, the act description must be a possible mode of self-determination. Whether through human imperfection, the laws of nature, or contingent

---

[16] See also, (Korsgaard, 2008, pp. 109), where she seems to conflate acting "under the idea of freedom" (autonomy) and being "free from any alien cause" (negative freedom) under the conception of the *free will*. As I discuss later, this conflates Kant's notions of the *pure rational will* and the *freedom of choice*.

[17] Complicating the matter further is the debate about whether we should frame Kant's notion of instrumental reason as a series of discrete hypothetical imperatives or one Hypothetical Imperative (Hill,1992, pp.17–37). Korsgaard is clearly some version of the latter view, but I will not address this debate further here.

[18] See also, "Skepticism about practical reason," in (Korsgaard, 1996a, pp. 311–334). For disagreement with Korsgaard's framing of the relation between "content skepticism" and "motivational skepticism," see (Brunero, 2004).

limitations on circumstance, there are some act type specifications that are impossible for sensibly affected rational beings. This covers both logical and real impossibility. For example, given restrictions on the logical structure of maxims, a rational being cannot will both P and not-P at the same time. Given restrictions on the real conditions of maxim formation, a sensibly affected rational being cannot will that his inclinations suddenly disappear, no matter whether he wants to be rid of them or not. These are only two examples, but they are sufficient to illustrate the point. Logically and really impossible maxims are not available to the rational being's faculty of choice.

The logical and real restrictions on maxim formation just outlined are only half the story. These restrictions merely elide the possibility of some maxims, while the restrictions' obverse allow for certain maxims as candidates for the faculty of choice. To be an account of freedom, if only negative freedom, there must be some domain of possible maxims left to the faculty of choice after the suitable qualifications made by the logical and real restrictions. To determine, in exactitude, the domain of possible maxims would overstep the remit of Requirement One, which seeks only to suggest that the will is free from full external determination. Therefore, Requirement One can say nothing about the domain of maxims available to choice other than the mere fact that the domain does exist and that it is conditioned by certain restrictions on possibility. This approach contrasts with a determinist view, which would take the domain of possible maxims to be determined wholly by what the laws of nature dictate. In contrast, Requirement One states that the laws of nature—included in the restriction on the real impossibility of maxims—condition the availability of maxims, but do not determine the choice of maxims. Requirement One leaves open a domain of possible maxims after defining the restrictions imposed by the sensible conditions of the natural world. This means that it presents a view of freedom, even if Requirement One alone tells us little about the domain of possibility itself. But this is all that negative freedom needs to say.

While Requirement One simply reiterates that there does exist a domain of maxims available to the faculty of choice, Requirement Two and Three specify what the domain looks like, what maxims are left to the faculty of choice, and how reason picks out certain features of these maxims as relevant. These two requirements outline the form that maxims must take if they are to be efficacious and considered as products of

the rational will's freedom. Requirement Two details the general form of efficacious acts, and Requirement Three details the notion of positive freedom.

Requirement Two corresponds to the restrictions on actions imposed by a hypothetical imperative. Coupled with the conditions of negative freedom outlined in Requirement One, Requirement Two shows the efficacious aspect of the *practical efficacy* view. Without getting unnecessarily caught in the debates that surround interpretations of Kant's notion of hypothetical imperatives, I will suggest hypothetical imperatives are contextually sensitive and amoral imperatives that both 1.) restrict mental states and 2.) suggest how to bring about a desired end.[19] With regard to the first point, hypothetical imperatives suggest that whoever wills the end also wills the means (IV, 417).[20] This is what is sometimes known as the transmission of one desire-like attitude to the belief in the means sufficient to bring about the attitude.

For instance, someone wants a sandwich, and this man believes that the means to making a sandwich are such that he needs bread. The desire for a sandwich transmits to the belief in the means. So now, if this person really wants a sandwich, he will also want to acquire bread. If we frame this point in terms of maxims, we can say that mental states must be in a fitting relation with each other if they are to be incorporated into a maxim. A maxim that expressed the desire for some end but also—barring any defeating considerations such as moral impermissibility—expressed an aversion to the sufficient means to this end would be irrational.[21] Likewise, a maxim that expressed the desire for some end but also expressed a desire for means that are not in proportion to the end would constitute an irrational response. Say that someone wants to eat a sandwich, and his maxim expresses the means of eating the bread's crust. Or, say that someone wants to eat a single sandwich, but his maxim expresses the means of eating a whole feast. These cases illustrate the importance of proportionality between means and ends. Notably, the maxims that conform to

[19] For one classic formalization of hypothetical imperatives, see P.S. Greenspan's (1975) "Conditional Oughts and Hypothetical Imperatives," where Greenspan argues that hypothetical imperatives are escapable.
[20] See (Timmermann, 2022,pp. 54) for a reading of the instrumental rule as chiefly about "wanting" the ends and means. This rendering of Kant's dictum as "wanting" is in direct contradiction to Korsgaard, who, in discussion of this point, states, "This distinguishes willing from mere wanting or wishing or desiring" (1996a, pp.94). I do not take this divergence to change the argument substantially.
[21] See (Korsgaard, 2009, pp.32) on bad or defective actions which fail to meet the constitutive standard.

hypothetical imperatives—in the sandwich case, for instance—start from the desire of some end. This is the context sensitive aspect of hypothetical imperatives.

By reading the above description of hypothetical imperatives, one would be forgiven for thinking that, in practice, all hypothetical imperatives achieve is a restriction on the relation between mental states.[22] Over and above the relation between mental states, hypothetical imperatives also suggest how a desired end is to be brought about. Kant is clear that hypothetical imperatives are *analytic* (IV, 417), in that reason determines the sufficient means already when it determines the desired end. The transition from merely wanting some end to thinking about the means of attaining that end opens the door for rationality to take command of inclination. Once the door is open for rationality to take command and suggest a possible maxim, then suddenly the rational will (*Wille*) can no longer forestall a judgement made on the proposed maxim.[23]

Here is where Requirement Three enters the picture. This requirement states that all maxims must conform to a law of reason, and thus, maxims cannot simply be a product of free choice (*freie Willkür*). In rough outline, Requirement Three posits the notion of positive freedom, or the freedom to will a maxim in accord with the law of pure practical reason.[24] Once reason is involved in the formation and enaction of a maxim, reason can further subject this maxim to a test of validity. But, maxims, by themselves, are subjective, and as such reason does not take them to have the force of necessity. If reason searches for a necessary justification of a candidate maxim, then it must be found elsewhere than the empirical incentive (IV, 408).[25] Later, Kant concludes that, if there is a supreme principle of practical reason, then it must be an *apodictic principle* that declares an action's necessity independent of its purpose (IV, 415).[26]

---

[22] The wide-scope vs narrow-scope debate about instrumental norms (Broome, 1999, 2020; Kolodny, 2005) is specifically about the coherence of mental states. But, as Fix (2020) argues, the Kantian constitutivist is concerned with functionally governed capacities, not rational coherence norms. In this case, the concern is about the capacity for efficient action.

[23] Compare with (Timmermann, 2022, pp.75–88).

[24] If the agent does not act *autonomously*, then the agent is acting *heteronomously* (IV, 433), which is determination by laws outside oneself. Therefore, an agent that attempts to escape his own self-determination is not free to choose arbitrarily what can be willed as a reason.

[25] In the cited passage (IV, 408), Kant argues that experience cannot furnish apodictic laws. Thus, experience is insufficient to provide a necessary justification for morality.

[26] In this passage (IV, 415), Kant distinguishes between the *apodictic*, and thus necessary nature, of a categorical imperative and the merely *problematic* or *assertoric* nature of hypothetical imperatives.

In *Groundwork* III, Kant declares that "even the most hardened scoundrel (IV, 454) will transfer "himself in thought into an order of things altogether different from that of his desires in the field of sensibility" (IV, 454).[27] When presenting the categorical imperative as a principle of autonomy in *Groundwork* II, Kant describes it as a synthetic a priori principle which commands apodictically (IV, 440). It is precisely this property of the rational will which allows for maxims that are chosen by virtue of their valid form, rather than the interest taken in the action's effect. It allows the hardened scoundrel to think himself as autonomous and judge maxims in accord with the moral law. The judgement passed with regard to a proposed maxim suggests that the action is to be brought about because it can coexist with the autonomy of the will, or in other words the action is judged *permissible.* Maxims which do not accord with autonomy of the will are *forbidden* (IV, 439). Further, we represent the will which subjects itself to its own laws, and which follows its obligations of duty, as a *dignified* will (IV, 440).

As I have listed them, these requirements are meant to get at the role maxims are meant to play in Kantian ethics. This is an important point to keep in mind when discussing the different types of rationalism that have been attributed to Korsgaard. Specifically, I will show that one of the main criticisms of Korsgaard's rationalism does not portray the SCC interpretation of *practical efficacy* in the most plausible way. Instead, I will choose to criticize Korsgaard's rationalism on more charitable grounds.

3.1 Distinctions in Moral Rationalisms

It is uncontroversial that Korsgaard's SCC theory is a form of moral rationalism. The view suggests that agents must subject their own motivations to rational introspection in order to decide what course of action is required by the situation at hand. It is the self-conscious construction of reasons and subsequent endorsement of these reasons that the SCC theory takes to be essential for an agent's response to normative requirements. If an agent is confronted with a problem, what can the agent

[27] The argumentative move at (IV, 454) involves the assumption that the hardened scoundrel is *accustomed to the use of reason* and will *wish* that he could act differently when confronted with examples of a *good will*. Hence, the wish itself is sufficient to transfer the scoundrel into the practical standpoint where he thinks himself as free of inclination.

take as a reason for action? And how can this agent make these reasons intelligible to others? It is what Korsgaard terms “the problem of the normative” (1996b, pp.93).

While the empiricist tradition is equally at home with the concept of a reason for action, the distinctively rationalist part of the SCC theory is the notion of self-conscious endorsement. But, it seems that much debate has been sparked about what exactly Korsgaard means by self-conscious reflection and endorsement. There is a passage from *The Sources of Normativity* which portrays the canonical view of the SCC theory:

> But we human animals turn our attention on to our perceptions and desires themselves, on to our own mental activities, and we are conscious *of* them. That is why we can think *about* them.
>
> And this sets us a problem no other animal has. It is the problem of the normative. For our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question. […] I desire and I find myself with a powerful impulse to act. […] Is this desire really a *reason* to act? (Korsgaard, 1996b, pp. 92–93).

From the passage above, it is clear that Korsgaard takes self-consciousness to be the ability to step back and submit one’s desires and beliefs to scrutiny. It is also clear that Korsgaard takes self-consciousness as a condition of the capacity to act for a reason, in contrast to the lower animals, which lack this capacity and merely act on first-order representations. What remains unclear is the relevant notion of stepping back from the object of thought and endorsing reasons to realize this object.

There are two prominent interpretations of Korsgaard’s rationalism.[28] Both views begin from the assumption that mental states are partially the product of passive receptivity. This is a cognitive process whereby the world impinges its data onto the mind of a conscious being. But, mere passivity is insufficient to ground anything so authoritative as the self-conscious activity of endorsing desires as reasons. In order to ground something like this notion of authoritative self-consciousness, one must provide an interpretation of the active self and its role in cognitive functioning. The views diverge in their description of the active self’s characteristic function and activity.

The first interpretation, which I will term the *Mental Objects* (MO) view, suggests that certain mental states like beliefs and desires are manipulable aspects of a rational

[28] See (Schafer, 2018; Smith, 2018) on different forms of moral rationalism and constitutivism.

self-constitution. This view treats mental states as equivalent to external objects in the sense that mental states may be moved about, repressed, or generated with sufficient mental labor. It is close to what Richard Moran (2001, 2022) has termed the Cartesian "internal theater" view of the mind, whereby one's mental states are luminous objects taking center stage under the view of the mind's eye. Under this view, mental states are so much cognitive furniture which one can manipulate in order to achieve internal organization. If, for example, the agent has luminous access to some irrational desire, then the MO view suggests that an act of mental will can easily change this desire such that the agent is no longer irrational.[29]

The second interpretation might be termed *Self-Authority* (SA). This view suggests that one must identify with certain of the mental states produced by the passive self. The activity of self-conscious reasoning is the ability to constitute oneself as a reasonable being, that is, an authoritative self which hangs over and above the merely passive self. The active self decides which mental states are reasons for action and identifies with these mental states, thus transforming them into reasons for action. Under this view, the reasons one takes up are the authoritative identity of the conscious being itself.

MO suggests that the active self starts by scrutinizing the passive self's mental states. If these mental states are found to be in an ill-fitting relation to the world, then it is the role of the active self to correct this discrepancy by fixing the mental states. In contrast, SA does not take the active self's role to be the general function of manipulating erroneous mental states; rather, SA assumes that the active self is the capacity to reason about which mental states one wants to identify with and endorse as a reason. Under this view, the active self is the true self, the self that is able to provide considered reasons which have been constructed from mental states. Those parts of one's identity that are endorsed—such as the habits one chooses to cultivate—are attributable to the choices of the active self. MO likens the activity of reasoning to the act of discarding some faulty object or producing a better object. SA, on the other hand,

[29] In framing the issue this way, I am largely drawing on Richard Moran's (2022) "Self-Consciousness and Self-Division in Moral Psychology," which criticizes interpretations like the MO View, though he does not broach the bootstrapping worry.

likens the activity of reasoning to taking on an identity, or molding oneself, insofar as possible, into the self that one would like to be.

As already shown above, the SCC model assumes the *practical efficacy* view. *Practical efficacy* is a view about the general conditions of forming a maxim. Under the SCC interpretation, Kant's incorporation thesis involves the incorporation of mental states into a maxim, such that this maxim can be treated as a universal reason. It seems, then, that the theory needs an account of the mechanism whereby mental states can be chosen or disregarded. Both MO and SA are views which, when supplemented to the SCC theory, can account for the agent's self-conscious engagement with his own mental states. It is my aim to show that MO cannot be accepted by SCC on pain of contradicting Requirement Three of *practical efficacy*. Specifically, acceptance of the MO view leads to a common criticism that the SCC theory justifies illicit bootstrapping because the free construction of reasons does not conform to a law of reason. If this is the case, then SA is the more plausible interpretation. Once I have shown this, I will turn again to the question of ambivalence. I will show that SA commits Korsgaard to proscribing ambivalence.

To see how the MO view might function in practice, I will turn again to Requirement Two of *practical efficacy*. Requirement Two states that the mental states incorporated into a maxim must have a fitting relation to one another. Let us bracket the issue of testing a maxim's moral worth for the moment, and think instead of a maxim that simply seeks to conform with a hypothetical imperative. A hypothetical imperative says that whoever wants the end must also want the means. In order to pursue some desired end, the belief in the means must be in a fitting relation with the desired end. If some person desired to quench his thirst, but he believed that the best way to do this was to jump in the River Tay, the relation between means and ends would be disproportionate and irrational. In this case, the means do not fit the person's desired end: implicit in a rational desire for quenched thirst is the specification that the means are in proportion to the person's desire. While jumping in the Tay would certainly end someone's thirst, it would also be a lethal act. Rationally desiring quenched thirst does not also entail a desire to die.

On this point, MO is sufficient as an account of the agent's self-conscious engagement with his own mental states. Once it comes to light that two mental states are in a disproportionate relation, the MO view easily accounts for the way in which the agent can update his mental states. One must simply notice that the relation between mental states is disproportionate, and then, following an act of 'mental will,' one can update one's belief or drop the desired end. Desires, on this view, are easily discarded, such that irrational mental states are only temporary roadblocks to full rationality. Under this view, no desires are brute, primitive, or immutable.

Although MO is sufficient to account for Requirement Two, this specific interpretation easily comes under criticism for modeling mental states as objects that can be willfully discarded. In other words, the MO view triggers the bootstrapping worry because it treats the construction of reasons as entirely free from constraints other than the constraint of simple means-ends coherence. For this reason, MO fails to provide a sufficient account of Requirement Three. Since this view cannot account for Requirement Three, I argue that it is not a charitable attribution to Korsgaard and should be rejected.

I will show how the MO view triggers the bootstrapping worry by suggesting an example of how this view treats cases of irrational desire. Suppose that Jones has just been to see his physician. The physician has informed Jones that he cannot drink any fluids for at least three hours after the visit. Only one hour has passed, but Jones feels parched. Suppose Jones sees a glass of cool, clear, thirst-quenching water before him. It would be natural for Jones, in this scenario, to respond to the perception of water by forming a desire to drink the water. But, Jones also believes that he should listen to his doctor and refrain from drinking. Intuitively, Jones should form a rational desire based on his belief that drinking the water is the worse than heeding his physician. Acting on the desire for water, in this case, is irrational.

Suppose, further, that Jones *rationalizes* his irrational desire to drink the water. He suggests to himself that enough time has passed, when in actuality it has not. He suggests to himself that the doctor's guidelines were not so authoritative, when in actuality they were. Suddenly, Jones believes that it is not so irrational to drink the water after all. From the assumption of the MO view, Jones can treat the desire for water as a

reason to drink because he has coupled the desire with a belief in proportionate and sufficient means. Suddenly, what was once considered an irrational desire has been bootstrapped into being a justified reason to take a drink. If the MO view is accepted, who would have authority to tell Jones that he is irrational? Jones has done the requisite mental labor to treat his once irrational desire as a reason for action. If no one else has access to Jones's mental states in the same way that Jones does, then it seems that Jones is the only one able to make claims about what reason he does or does not have.

While this specific bootstrapping worry is triggered by the MO view, I take it that MO actually contradicts Requirement Three of the *practical efficacy* view. Recall that Requirement Three suggests that an agent's maxims must conform to a law of reason. If we frame this requirement in terms of practical reasons, then it would dictate that any arbitrary mental state, barring some mark of relevance, cannot be treated as the best reason one has. In addition, simply treating a mental state as the best reason one *could possibly* have does not necessarily make it the case. Of course, if agents were not cognitively limited beings, then perhaps they would always know what would be the best reason that could possibly be constructed. But, the fact that agents cannot know counterfactually which reason would be best does not mean that any arbitrary mental state can be treated as a justified reason. Requirement Three sets restraints on how mental states can be conceived as reasons and which reasons are justifiably treated as *good reasons*. Therefore, MO fails to provide a sufficient account of Requirement Three.

In contrast, the SA view is sufficient to account for Requirement Three. To see that this is the case, one should look to the notion of endorsing one's mental states. What would it mean, for example, to endorse a desire and to identify oneself with the desire that one has? If a desire, in the general sense, is a pro-attitude that pushes an agent toward achieving the object desired, then endorsing a desire would also act as an endorsement of the thing desired as *that object which is to be brought about*. SA does not say that one can pick one's desires at random. It says that, of the desires that one has, one must choose to endorse some desires over others. The desires chosen are then treated as expressions of the agent's true self. To endorse some desire, then, is to endorse an object under the description of *that which is to be brought about because it is the object of some pro-attitude which I have chosen as indicative of my true self*.

If we return to the case of Jones, we have a scenario in which there are two conflicting desires. One is a desire for water, and the other is presumably a desire for a healthy or prudential state. Unlike MO, the SA view suggests that the agent is not only reasoning about his own mental states, but about the objects of the mental states.[30] Contrary to bootstrapping an irrational desire into being the best reason that one has, this view does not deny that there are certain features of the world which condition what could possibly be endorsed as a good reason. In other words, rationalizing the desire to drink the water, and endorsing this desire, cannot make it the case that other, possibly better reasons are invalidated. Ultimately, what an agent could take to be most rational will depend upon the SCC theory's account of valuing, which I will turn to in the next chapter. For now, suffice it to say that SA does accord with Requirement Three because rationally justified desire endorsement is not the same as arbitrarily choosing which desire one wants to fulfill the most. There is a fact of the matter about which possible desire would be most rational to endorse. Whether the agent has the cognitive wherewithal to recognize what it would be most rational to endorse is another question entirely.

Under the SA view, if an agent desires health, recognizes that this is the most rational desire in his desire set, and yet the agent endorses the desire to drink the glass of water instead, this does not change the fact that health is the most rational desire in this scenario. Unlike MO, which treats mental states like a black box separated from the world, SA can account for cases of practical failure. The SA view recognizes that some desires are inveterate and cannot simply be willed out of the agent's mind. Sometimes these desires are strong and the agent is weak. Requirement Three does not have to claim that agents always are rational; rather, it only needs to claim that the agent's possible reasons depend upon the mental states that the agent has. This precludes rationalizing oneself into constructing an irrational mental state into a good reason. Therefore, where MO contradicts Requirement Three, SA does not. The SCC theory should not be interpreted as subscribing to the MO view.

[30] This is sometimes called "transparency"; see (Kolodny, 2005; Korsgaard, 1996b, pp.17; Moran, 2001,pp.60–65). For criticism of Korsgaard's transparency condition, see (Copp, 1997).

Ultimately, I will suggest that SA still results in a type of bootstrapping worry, but this new bootstrapping worry is subtly distinct from the one that has been suggested above. In particular, this new bootstrapping worry results from cases of moral reasons defeating non-moral reasons. If the SCC theory assumes that every agent has a desire to perform the moral action, and if it is also assumed that the moral action is always the most rational action, then I will argue that the theory must model rational agents as wholeheartedly endorsing moral reasons which easily defeat non-moral reasons.

I will suggest that, for the SCC theory, all reasons are homogenous in two ways. First, all reasons start with the same content. Remember that, in the process of deliberation, all reasons start as desires of inclination before they are crafted into proper reasons. Second, if all practical reasons contribute to the same function—where some reasons are simply judged better at this function than others—then the reason which contributes to the function in the best way defeats the other reasons. In the case of the SCC theory, practical reasons all share the function of determining one's agency. From point one and two, it follows that one should construct the reason that contributes in the best way to the function of agency, and any mental state which does not contribute to this proper functioning simply cannot function as a good reason in the proper sense.

The view suggested above is a theory of *moral virtue*, in which the function of practical reasons is to constitute the agent as fully rational. This view leads to a case of bootstrapping which differs from the other bootstrapping case because it relies on the concept of *moral defeaters*. In the former bootstrapping case, the MO view suggested that an agent could bootstrap bad reasons into being good reasons. In the new bootstrapping case, the SA view suggests that there are certain objective, though internal, reasons that easily defeat all other possible reasons because these reasons contribute best to the virtuous function of agency. Therefore, in the new bootstrapping case, it is not clear how an agent can both recognize the moral reason and still desire to do a non-moral act. Ultimately, I will argue that because of these assumptions the SCC theory cannot account for cases of mental ambivalence. Before turning to that point, though, I will look at how the SCC theory is committed to the wholeheartedness view.

3.2 Undermining Oneself Wholeheartedly

Since the most plausible interpretation of the SCC model has been identified, we can now see how this view commits Korsgaard to some form of the wholeheartedness view in moral psychology. As stated above, SA suggests that agents endorse certain mental states for the purpose of constructing them into reasons. In terms of belief, a mental state which tracks an independent world, the agent endorses his representation as an accurate depiction of reality. That the agent seeks to represent the world correctly gives the agent *epistemic reason*. In terms of desire, a mental state that represents some object to be brought about, the agent endorses his representation as something that really should be brought about. The representation of an object that should be brought about gives the agent *practical reason* to perform the relevant action that would bring about said object.

When engaging in the activity of practical reasoning, Requirement Three of the *practical efficacy* view suggests that some objects must be represented as all things considered better objects. This is sometimes described as a reason functioning as a defeater, in the sense that it defeats competing reasons. If the SCC theory is committed to some version of the wholeheartedness view in moral psychology, then the theory is also committed to framing moral reasons as defeaters. This, of course, leaves open an important question: what is the function of a practical reason such that moral reasons defeat non-moral reasons? The answer to this question will largely shape what the relevant notion of defeat in this context will look like. For the SCC theory, the function of a practical reason is to constitute or determine the agent into being a rational self. Korsgaard terms this view the “Constitutional Model” of the rational will (2008, pp.100–126). If the scope of moral self-determination includes one’s mental states—and I will argue that, for Korsgaard, it does—then the agent must constitute himself wholeheartedly and undermine his own mental states in cases of moral defeat. This is a specific interpretation of *moral virtue*, and it depends upon the “Constitutional Model.” However, before I turn to that issue, I will address how Korsgaard commits herself to the wholeheartedness view in moral psychology.

The SA view suggests that agents must identify with the reasons they have for action. If an agent finds that he cannot identify with the desire he has—that is, an agent finds that he cannot endorse some desire as an adequate candidate for a reason—then

the agent must have a desire to perform some other action. When an agent does not wholeheartedly endorse some desire picked out for reason candidacy, then this is a sign that the agent must submit his desires to further self-conscious scrutiny.

An analysis of the concept *wholeheartedness* suggests that it is a mode of desire endorsement. Wholeheartedness is an act of desire endorsement such that the agent truly believes the endorsed desire is the best desire. Anything less than full belief or full desire would not be *whole*hearted. If the desire one endorses cannot be willed wholeheartedly as a reason, then the agent is not fully rational. Only wholehearted willing could be a sign that the agent takes himself to be fully rational. By contrast, wholeheartedness is a sign that the agent has a reason, endorses it as the best reason he could have, and believes himself to be fully rational by virtue of this endorsement. Of course, there is room, under this view, for cognitive limitations and ignorance. Just because an agent is wholehearted about his reasons does not mean the agent is not mistaken about the best action. However, ambivalence is a sure sign that the agent is doubtful of his reasons. It is a sign that the agent is aware his reasons do not stand up to scrutiny. Thus, it is a sign that his reasons should be probed further.

Given that ambivalence about some reason is a sign that one should subject one's reasons to further scrutiny, the SCC model must proscribe ambivalence if it takes reasons for action to function as possible determinations of rational agency. If an agent is ambivalent about his reasons, it is likely that, on due reflection, he will discover that he could have willed a better reason. It is an assumption of the SCC model that the Categorical Imperative is a principle which constitutes the function of rational agency. Thus, on due reflection, the agent can at least discover that his reason does not pass the test of universalization. The agent's heretofore ambivalence showed a tacit awareness of the Categorical Imperative. The rational response, in this instance, is to step back from one's desires, inspect them, and then legislate reasons which one can will wholeheartedly. This shows that Korsgaard's SCC model, in this respect at least, is like the other Self-Constitution theories. An agent who knowingly produces a good effect from an action with poor justification is not acting from the moral law.

I have just shown that Korsgaard's SCC theory does subscribe to a version of the wholeheartedness view. I will now suggest that this interpretation of moral motivation is

not extensionally adequate to account for the psychology of ambivalence. Specifically, I will suggest that the SCC account flattens all mental states into possible content to be plugged into a single function, namely the function of rational agency. In particular, once all mental states are considered as candidates for the legislation of practical reasons, then the theory suggests that certain mental states—such as an agent's ambivalence about a situation—should be discarded because the agent has defeating moral reasons.

Before I show how moral defeat creates the new bootstrapping worry, I will address how it is that the SCC theory takes the scope of self-determination to include one's own mental states. This view largely depends on the "Constitutional Model" of the rational will, which Korsgaard sets in opposition to the "Combat Model" of the will (2008, pp.100–102). The Combat Model is a view of agency that suggests actions are produced when some internal combat, such as the combat between reason and passion, leads to the strongest force winning and determining the agent. By contrast, the Constitutional Model suggests that agents fully constitute themselves through actions that can be attributed to the agent as a *whole* person. Furthermore, under the Constitutional Model, "What makes an action bad," Korsgaard writes, "is that it springs in part not from the person but from something at work *in* or *on* the person, something that threatens her volitional unity" (2008, pp.102). Clearly, then, the scope of rational self-constitution includes the agent's mental states because, on this view, the agent must set all his desires and beliefs in order, such that the agent is functioning virtuously.

The fact that the SCC theory's chosen analogy is a political constitution means that irrational mental states can be left to their own devices so long as they do not interfere with the rational executive.[31] Of course, like a dissident that threatens the safety of the rational leader, irrational mental states are targeted with force to secure the function of rational agency if and when leadership is threatened. It is not that conflict between rationality and inclination is missing under the Constitutional Model. A constitution, for example, sets the limits within which different interest groups can justifiably compete. However, it is the case that, under this view, actions *should* be determined by the formal restraints set by rational agency, rather than the natural

[31] This is clearest in Korsgaard's attribution of the Constitutional Model to Plato's *Republic* at (2008, pp.102–109).

combat of inclination. The worry is that the Constitutional Model, as it has been described, goes too far in its emphasis of rational self-determination. Moral virtue and the prospect of moral defeaters should not threaten the possibility of modeling an agent who is both moral and in a state of mental ambivalence. While it is plausible that the Constitutional Model is meant to allow inclinations the requisite room to operate within the constraints set by the rational agent, I argue that the space open to inclinations, when the agent faces a moral duty, is inadequate.

To see that this is the case, one can focus on Korsgaard's reading of Kant's Friend of Humanity passage in *Groundwork* I. The Friend of Humanity is an example that Kant gives in order to illustrate why an action merely in accord with duty, but not from duty, lacks moral value (IV, 398). The example is this: there is a man who persistently acts beneficently when confronted with others in need. However, the man acts from his own sympathetic feelings, not for the sake of moral action itself. But, counterfactually, given the conditions of the man's sympathetic nature, both act types—the action from duty and the action merely in accord with duty—have the same act token. Empirically, there is no way to differentiate these actions. The true method of differentiation is the sympathetic man's maxim. Given that the correct conditions hold, a maxim to act from the incentive of sympathy only produces an act of beneficence generally, insufficient for Kant's project in *Groundwork* I of analyzing the concept of duty. Furthermore, and relatedly, an action done from sympathy cannot account for the notion of a morally good action.

The insufficiency of an incentive of inclination merely in accord with duty is not a conclusion that Kant asks his reader simply to take on faith. He shows that this conclusion holds by asking his reader to suppose that a terrible tragedy befalls the sympathetic man. The tragedy is such that the man no longer has the incentive to be sympathetic. The joy with which he used to help others is gone, replaced with nothing but apathy. But the duty remains; the man should act with beneficence. Counterfactually, given these conditions, the reader can finally apprehend the empirical difference in such a scenario. The different act types produce divergent act tokens. If the man were to act from the incentive of his own apathetic inclinations, he would shirk his duty. If the man acts from the incentive of respect for the law, he will perform his duty

and help those in need, despite being inclined to act otherwise. The clarity that this case sheds on the question of moral motivation is precisely why Kant often discusses cases where inclination and duty diverge. The tragedy of the sympathetic man reveals that inclinations are subject to changes in circumstance, and thus the determination of will provided by inclination is weak and lacks the necessity of law.

The above interpretation of this passage is fairly uncontroversial, but I will focus now on Korsgaard's interpretation of motivation. In particular, I will argue that Korsgaard's interpretative assumptions fail to capture what is distinctive about this case: the man is *both* apathetic and moral. It is not that the man suddenly sets his apathy aside in order to act from the moral law. On the contrary, what makes the case distinctive is that the man acts morally *despite* his inclination to do otherwise. I will argue that the SCC theory fails to capture this key feature of the case because its presentation of moral motivation relies on the Constitutional Model and is too strongly rationalist. The problem arises because the SCC theory interprets the categorical imperative as a decision procedure which subjects a homogenous set of mental states to a universalization test.[32]

With regard to the naturally sympathetic person and the question of moral motivation, Korsgaard writes, "Kant's thought is that a reflective person asks herself whether the consideration on which she proposes to act may really be treated as a *reason* to act" (Korsgaard, 2008, pp.186). She goes on to describe the naturally sympathetic man applying the dual aspect theory of motivation.

But two of the claims that Korsgaard makes are anything but clear from the *Groundwork* passage in question. The first questionable claim is about the relevant sense of reflecting upon one's maxims. It is not clear that Kant believes the sympathetic person beset by tragedy has reflected upon his maxim and considered whether the act and its purpose is *good* in the relevant sense. What is clear is that this hypothetical person does a good deed, and he does it without inclination telling him to do so. But doing an action from the moral law does not necessarily mean that a person has subjected his maxims to considered thought. It might very well be clear, when confronted by a situation that requires beneficence, what the correct thing to do would

[32] See "Kant's Formula of Universal Law" in (Korsgaard, 1996a, pp.77–105), esp. pp. 92–94.

be, and this without very much by way of reflection on reasons. Given Kant's insistence that the pre-theoretical notion of morality is correct, and his insistence that *Groundwork* I is an analysis of our everyday moral concepts (IV, 397), it is unlikely that this model of reflective scrutiny, whereby the agent determines reasons for action through considered thought, is the one that Kant has in mind.

The second questionable claim is more tacit, but I believe that it is clear enough from the above passage if one returns focus to the dual aspect theory of motivation. Remember that the SCC theory takes motivation to consist in two aspects: the incentive and the higher-order principle of volition. Under this interpretation, then, the sympathetic man beset by tragedy inspects his possible maxims and asks himself whether any can be willed in accord with the principle. But, while one possible reason fails the relevant deliberation procedure, both possible reasons are not really distinct. Both maxims have content by virtue of the incentives suggested by the world. And the possibility of either maxim being treated as a reason comes from the capacity for reflective deliberation.

Both maxims are reasons for action, and the maxim which passes the higher-order deliberation procedure becomes the all things considered best reason for action. However, there is a major problem in framing the issue of motivation in this way: if there is no mark to distinguish between practical reasons—that is, all possible practical reasons have been homogenized as determinations of rational agency—then moral reasons cleanly defeat other practical reasons when these reasons diverge in their suggested determination. Since it is assumed that the scope of practical reasons includes the agent's mental states, then it is unclear how an agent can be mentally ambivalent in moral cases.

If we take the man beset by tragedy as an example, the inclination he has to act without beneficence was once a candidate reason of the same kind as all other mental states, including the possible desire to act with beneficence. But, once the categorical imperative suggests which maxim serves as the all things considered best reason, how could this man cling to an emotional state which conflicts with the rational function of virtuous agency? Rationality does not seem to permit it. And this, even when the man *does* perform the rationally required action. If the man's emotion is itself treated as a possible practical reason *of the same kind and purpose* as moral reasons, then how can

this man retain his complex emotional state in the face of rationality's ineluctable march?

This model of motivation suggests that the man, once he has sufficiently reflected, should simply bootstrap himself out of having a contrary to duty inclination. But it is questionable whether this is really the remit of rationality, and whether this is something that we want our theory of practical reasoning to do. So long as our theory of practical reason permits mental ambivalence while proscribing motivational ambivalence, then we will have a theory with the extensional adequacy to model cases in which the agent both performs the rational action and feels conflicted. The SCC theory, as I have presented it here, does not currently have the wherewithal to deal with these cases.

4. Conclusion

While I do not want to deny that Kantian virtue has an affective aspect, I have argued that the constraints imposed by rationality are too strong if we model them as easily defeating cases of mental ambivalence. As Alix Cohen (2017, 2018) has argued, virtuous willing cultivates certain dispositional states that both follow from and enhance the virtue of the rational agent. In the *Second Critique*, Kant writes that there is a negative feeling associated with acting from duty: "*Contentment with oneself* [*Selbstzufriedenheit*] must necessarily accompany consciousness of virtue" (V, 117). And further, he notes a positive feeling "which cannot be called happiness […] nor is it, strictly speaking, *beatitude*, […] but it nevertheless resembles the latter" (V, 119). Still, earlier in *the Second Critique*, Kant also criticizes the stoic and rationalist-Wolffian principle of *internal practical perfection* as "nothing other than *talent* and what strengthens or completes this, *skill*" (V, 40-41).

It seems, then, that a Kantian theory of moral emotion must thread the needle between these two extremes. I have not said much about what the theory should look like. Instead, I have focused on one model that I take to be insufficient. If we turn back to the Hamlet case, it is easy to see what I mean. Suppose that Hamlet decides to spare Claudius, not because he is motivationally ambivalent, but because he decides that this would be an immoral action. Is it really rational for Hamlet to stop hating

Claudius, the man who has married Hamlet's mother after murdering his father? If Kantianism provides a theory of moral emotion distinct from the internal perfection of the stoic, then it seems that we should say that Hamlet is rational, in this case, despite his mental ambivalence. As I have shown, the SCC theory is unable to meet this charge.

Chapter Three

1. Introduction

There are several salient questions about the assumptions at play in the SCC theory. Most importantly, we should ask what notion of valuing the SCC theory puts forth such that there is an account of the comparative and superlative quality of reasons. In Kantian philosophy, there are two prominent accounts of value. First, there is the *relationist* account, to which the SCC theory subscribes. Relationism is the view that value emerges from the relation between things. Second, there is the *substantialist* account, to which philosophers like Rae Langton (2007) and Allison Hills (2008) subscribe. Substantialism is typically the view that Kantian ethics rests on some axiological foundation, such as the value inherent in autonomy, rational nature, or humanity. Ultimately, I will say very little about substantialism, its interpretative viability, or its philosophical method.[33] Instead, I will focus on the SCC theory's account of relationism and some of the philosophical questions that arise from its account of rational choice.

In this chapter, I investigate the conditions which allow for both *moral* and *non-moral value* under the SCC theory. In particular, this chapter concerns the conditions under which a first-personal notion of goodness, both moral and non-moral, for one agent can be intelligible across the domain of all rational agents. And in turn, I will investigate how this account hopes to connect the first-personal conception of the good to a concept of goodness for all.

In order to begin this investigation, I will turn first to certain questions concerning the SCC theory's notion of rational choice. First, how can an agent rationally choose the constraints he imposes on his own activity of rational choice? And, second, is the SCC theory's account of rational choice truly a universal activity? The first question impugns the viability of thoroughgoing constructivism as an account of normativity. The *ex ante* choice of rational constraints which are imposed onto the agent's own ability to make rational choices *ex post* cannot rely on the same notion of rational choice. Metaphysically speaking, it is difficult to explain how something might impose

[33] For criticism of Kantian substantialism, see (Bader, 2023; Fix, 2023; Sensen, 2009, 2022).

constraints onto itself without these constraints being arbitrary guidelines broken at will. This first question has been a perdurable criticism of Kantian philosophy's notion of autonomy as the rational will's self-imposed constraints. The explanatory task for SCC theorists is slightly different from that imposed on classical Kantians, given that SCC theorists emphasize agency rather than transcendental idealism. The second question, on the other hand, asks for a philosophical justification of the SCC theory's specific notion of rational choice.

Much of this chapter will attempt to show how the SCC theory plans to meet these questions. As such, much of the chapter is expository rather than argumentative. Still, there are several passages in this chapter where I raise concerns about the SCC theory's framing. Through sections 2.1-2.3 below, I aim to present the SCC theory's account of agency and valuing. Through sections 3.1-3.2, I aim to answer the two questions raised above by showing how the SCC theory both explains and justifies its notion of agency.

2.1 Two Distinctions

The *relationist* strategy in metaethics is a philosophical commitment that suggests the concept 'good-for' takes primacy over the singular concept of 'the good.' The view has a strong pedigree, including expressivists such as RM Hare (1952/1978, pp.127-150) and constructivists such as John Rawls (1971/1999, pp.350–351).[34] The simplest way to characterize the view is that the term 'good,' when used in discourse about value, typically means that the thing described as good is *good-for* some purpose, act, state of affairs, or object. This is a functional use of the term 'good' which is meant to precede any justifiable discourse about 'the good' in abstracta, the 'final good' as the ultimate *telos*, or the 'highest good' in the hierarchy of value. In other words, discourse about the goodness of something always takes a stance on the relation of the thing's goodness to some other thing. If someone expresses the proposition *that food is good*, he is declaring that food is good *for him*.

[34] Notice, that (Rawls, 1971/1999, pp.351) gives three relational notions: I.) *X is good if it is good at the function of Xs*. II.) *A is a good X for K, where K is a person*. III.) *A is a good X for K given Ks plan of life*.

Several Kantians have suggested that Immanuel Kant himself subscribed to some form of relational theory of the good. This interpretation of Kant often emphasizes the notion of rational choice. Value, under this interpretation, is a product of what an agent *chooses* to regard as valuable. Among the passages cited as evidence of this view is the first line of the *Groundwork*, where Kant declares, "It is impossible to think of anything at all in the world, or indeed even beyond it, that could be taken to be good without limitation, except a GOOD WILL" (IV, 393). Here, the relationists take Kant's emphasis to be agency: the good will is the capacity to act, choose, and reason practically, while the good is always *good for* some agent. If the good will is the only thing thinkable as good without limitation, then the value of everything else is conditioned by the contribution it makes to the agent. Kant does come close to saying this when he suggests that even the good of happiness, a good that all accept merely by being sensibly conditioned rational beings, can only be considered truly good when it is enjoyed by someone with a good will (IV, 393; V, 25-26, 62).

In a now famous essay titled "Two Distinctions in Goodness" (1996a, pp. 249–274), Korsgaard focuses on the goodness ascribable to states of affairs, acts, and purposes as objects of choice. In the essay, Korsgaard suggests that philosophers too readily equate notions of goodness that should be kept distinct. Failing to keep these notions separate prejudges the possible contours of our theory. This equivocation unduly restricts the philosophical theories that can justifiably be taken up. The two distinctions can be set forth in the following schema:

Intrinsic/ Extrinsic Distinction	**Means/ Ends Distinction**
Intrinsic – something has value intrinsically when value is inherent.	*Final Value* – something is valued as an end when it is valued finally, without the purpose of achieving some other end.
Extrinsic – something has value when value is imparted from an external source.	*Instrumental Value* – something is valued as a means when it is valued for the purpose of achieving something else.

Korsgaard's main claim in the "Two Distinctions" essay is that philosophers have too often conflated intrinsic value with final value, and this has led to a philosophical oversight which disregards the Kantian theory of value (1996a, pp.250–253).[35]

Foremost among the concerns raised in the "Two Distinctions" essay is the philosophical effect of conflating final value with intrinsic value. This view forecloses the possibility of something's value coming from the mental states of an external valuer. For example, if final value and intrinsic value are conflated, then something valued not as a means, but for its own sake—such as the beauty of a painting—must either have some objective value which inheres in the beautiful object. Or, the value must be inherent to the mental state of perceiving the beautiful painting. In the first case, value is a real object in the world and independent of the human disposition to take these qualities as valuable. In the second case, value is simply a subjective mental state, such as pleasure. Neither of these two cases can accommodate the suggestion that objects are really valuable *because* a valuer does value them. Instead, value is located squarely in one place or the other, either the external world or mental states. The relationist, by contrast, would reject this as a false dilemma.

By separating intrinsic value from final value, we can raise the possibility that a thing is valuable finally but extrinsically. In other words, the value in this case comes from an external source, but the object is valued for its own sake. For example, a painting in this instance can be said to be valued for its own sake, and also the painting can be said to be valuable because some valuer *does value it*, where this means that the valuer has the correct evaluative mental states directed at the object. This is why the relationist often emphasizes the phenomenon of rational choice. Value, under this interpretation, is a relation between some object and the evaluative attitudes taken toward said object. Therefore, what it means for something to be valuable is for this thing to be choiceworthy under certain specified conditions. What those conditions are varies depending upon the thing valued. Art is valuable under conditions in which the agent has a desire to view something deemed beautiful, thought-provoking, or some

[35] But, in contrast, see (Langton, 2007), esp. pp. 162–165 for an alternative presentation of the distinctions, which adds space for *something being valued instrumentally* for its effects and *something having instrumental value* as a means to effects.

other specification of relevance. Reciprocity is mutually valuable under conditions in which multiple agents need to coexist. Something valued universally by all rational agents, then, is simply something that all rational agents with the capacity for evaluative judgement would choose under the right conditions. Of course, the task of listing the correct conditions which all agents universally share is near Herculean in its philosophical difficulty.

The activity of rational choice itself is at least a plausible candidate for the conditions which all agents universally share. For one, once we assume the first-personal perspective, rational choice is an activity that all agents do take part in. From the first-person perspective, rational beings *must* reflect and act. To do otherwise would be to deny authority to one's own first-personal experience, choosing instead to cede authority over one's own thoughts and actions (Korsgaard, 2008, pp.114). In addition, there are constraints on rational choice, such as mutual benefit, that it is easy to model on merely instrumentalist interpretation of rationality.[36] While these two points provide plausibility to the idea that rational choice is a universal condition, plausibility is not sufficient to purchase theoretical justification.

2.2 'Good' as Functional Term

Foremost among the matters of theoretical concern here are questions about the relevant notion of rational choice. How should we analyze the concept of rational choice? And how is it that we can provide a first-personal description and analysis of rational choice, such that it is justifiably universal? This section seeks to provide some answer to these questions.

If we bracket the question of justification and assume for the moment that the activity of rational choice is a universal activity that all do take part in, then we can analyze the concept of a *rational choice* to see what this concept might yield. At the very least, the concept requires that the rational choice be directed at some object, action, or state of affairs. Rational choice is intentional and must be directed at something. In order to direct the intention of rational choice, there must be an *evaluative judgement* that picks out certain things as objects worthy of choice. The evaluative judgement

[36] See, "Two-Person Cooperative Games" in (Luce and Raiffa, 1957, pp.114–154) for a classic treatment.

declares that something is good, and thus, worthy of being brought about. This is what some philosophers describe as the functional nature of the term *good*.

One might raise a worry here about the framing of rational choice and choiceworthiness. We have just analyzed *goodness* as a relational concept which refers to something rationally chosen under the right conditions. And in order to explain what the right conditions are, we have appealed to a notion of *rational choice* which suggests that rational choice directs itself at objects judged to be good. Surely, this is circular reasoning. We cannot claim that goodness is what the agent rationally chooses, and then explain rational choice by appealing to goodness.

In order to dispel the circle, or at least render it less viscous, we can look to cases of evaluative judgement. Suppose that a friend expresses a desire to eat ultra-processed cakes everyday of his life. And, suppose further that the friend expresses the desire for ultra-processed cakes, in the form of a hypothetical imperative, as a means to living a healthy life. Is the predilection for ultra-processed cake truly a good means to the desired end in this case? The friend has made a judgement, and this judgement suggests that the object of his desire, the habitual eating of ultra-processed cakes, is something that is to be brought about. Thus, the desire is endorsed as a reason for action. However, even though the friend has expressed that he thinks the object of his desire is choiceworthy, it is not necessarily the case that ultra-processed cakes truly are the most rational object of desire in this scenario. As I argued in the last chapter, the SCC theory does admit of good, bad, better, and worse reasons for action. In this scenario, one could try to convince the friend that, in actuality, eating ultra-processed cakes as a daily habit is unhealthy. Thus, one could convince the friend that he has made an erroneous evaluative judgement.

Notice, though, that the above claim can be made without appeal to anything like some inherent value property, such as non-moral goodness, that the ultra-processed cake fails to instantiate. The point of the argument is that the ultra-processed cake habit is not *good for* the friend. The goodness or badness of the ultra-processed cake is entirely dependent upon the effect that it has on the friend. Without the friend in view, the ultra-processed cake is neither good nor bad but simply a cake with large quantities of preservatives and high-fructose corn syrup. In effect, the SCC theory will have to say

that, given the type of being that the agent is, there are certain features of the world which it would be *most rational* to desire. Furthermore, given the type of being that the agent is, there are certain intentions it would be *most rational* to endorse. This leaves open two important questions which I hope to answer in the next section. One, what type of being is the agent? And, two, what is the relevant notion of rationality?

Now, the case of the friend is obviously not a case of moral judgement but simply a case of prudential judgement. Therefore, we are talking about a case where some object has been judged non-morally good. Some will argue, rightfully, that it is difficult to generalize the conclusion found in the case of prudential judgements to cases of moral judgements. In the next section, I will spell out the conditions which allow for correct and incorrect moral judgements. In particular, moral judgements will depend upon what the SCC theory takes to be the *final good*. Before I get to that argument, though, I will return to analyzing the notion of evaluative judgements and the functional use of the term good. I take it that the foregoing has been sufficient at least to show that sometimes evaluative judgements can be mistaken even when we are speaking about the *good for* some agent. And it follows, then, that it is not merely circular to speak of something being *rationally chosen* and to explain this by appealing to the object's *choiceworthiness*.

There is a two-fold sense in which an evaluative judgement might be described as functional. First, evaluative judgements of a thing's goodness are meant to draw attention to the choiceworthiness of that thing. The function of the term *good* is to recommend something as worthy of choice. But this is not distinct from what would be said by noncognitivist theorists. In fact, prescriptivism as a moral theory suggests that prescribing choice direction is the main function of normative discourse (Hare, 1952/1978). There must be some feature, over and above the function of prescription, which makes evaluative judgements functional in a rationalist sense.[37]

Focusing on the rationalist features of Korsgaard's SCC theory brings us to the second sense of an evaluative judgement's functional nature. This focus allows us to distinguish the SCC theory from noncognitivist prescriptivism. Over and above the mere prescription of something's choiceworthiness, the SCC theory also takes it that objects

[37] (Sensen, 2022) argues that Kant's account of non-moral value is a form of prescriptivism.

and actions have intelligible functions ascribable to them by virtue of the purposes and intentions of a rational agent. For example, a human-made artefact such as a knife has the function of cutting, slashing, or stabbing because this is the rational intention behind the knife's invention and use.

Notably, the ascription and representation of some non-rational object's function hinges on the rational agent's own functionally structured capacities and the interaction of the rational agent with the world. For example, there are certain features of a sharp object—such as the very fact that the object has a sharp edge—not freely chosen by agents, that allow the object to be conceived as a tool. Suppose the agent is confronted with a sharp object and must determine what is to be done with this object. Why should the agent attribute the knife-function to the sharp object, rather than, say, the function of a decorative art installation? There are several reasons why the choice of function might be made. The sharp object might be more effective as a tool than as a source of aesthetic pleasure. Given the agent's role in this scenario as the ultimate arbiter of which reasons are to be most relevant, the attribution of an intelligible function to some object is also a matter of rational choice, to the extent that the function is chosen based on rational considerations. The rational choice of intelligible functions might be determined in part by objective features of the world, in part by the history of socially directed choice, and in part by individual choice. This is a rationalist picture not because the SCC theorist assumes that reasons can be provided for one's intentions, which surely is something the SCC theorist believes. It is a rationalist picture because the SCC theorist assumes that functions are fundamentally intelligible and shareable aspects imposed on the world by agents who share the same rationally structured cognitive capacity for choice.

If agents are free to choose which intelligible functions apply to artefacts, then how do the choices of various agents converge? What factor precludes the rational choice of intelligible functions from being completely anarchic or contingent? SCC theorists provide a response to this question by appealing to a teleologically structured account of reasoning. The theory posits a conceptual connection between evaluative judgements and the concept of the final good (Korsgaard, 2013). Importantly, the conceptual connection at issue here is unlike that alleged by theories which attempt to

model the final good as a simple aggregate of preferences revealed by a *homo economicus*. That theory too would provide an account of conceptual connection, but this conceptual connection would be so tight as to render evaluative goods and the final good only quantitatively distinct. A simple aggregate of evaluative goods does not constitute the final good. The SCC theory suggests a conceptual connection between final goods and evaluative goods that is closer to Aristotle's unmoved mover, the telos that imparts functions onto material reality. But, in contradistinction to Aristotle's view, the final good is not independent of the agent's own rational function.[38]

By contrast, the SCC theory suggests a view whereby agents' choices converge on functionally governed descriptions of things. The convergence of disparate agents' choices is explained by the structure of rational agency, which is governed by a universal, normative principle. From this principle, all agents do share an end that is representative of their characteristic function as self-conscious beings. The intelligible function of agency is a controversial issue which I will return to below. But, before I describe this philosophical commitment in detail, one can see why the intelligible function of agency is a necessary assumption for the SCC theorist's argument. If SCC theorists want to argue that universal evaluative judgements arise from the function of agency, then they must show how exercising this capacity for agency leads to an end that all must share by virtue of being such agents.

One can see why the intelligible function of agency is a necessary component of the theory by turning to conflict of interest cases. This kind of case proposes a scenario in which the interests of various agents coincide contingently. Kant suggests that there are two notions of happiness, the idea and the state of being happy. The idea is an objective, though generalized, notion common to all agents. But, being in a state of happiness is a more determinate, though subjective, notion of happiness formed from the first-personal standpoint of the deliberator (V, 430).[39] Given this theoretical assumption, all agree on the first conception of happiness by definition. But, if we assume there is a scenario in which multiple agents agree on the first-personal notion of happiness, then this is simply a contingent case of intersubjective agreement. In the

[38] See Korsgaard's "Aristotle and Kant on the Source of Value" in (1996a, pp.225–248).
[39] See also (Timmermann, 2022) pp. 15–16.

case of contingent agreement, harmony between agents could be undermined by a simple change in circumstance. The interests of multiple agents coinciding on a scarce good, for instance, would surely undermine the state of equilibrium.

Assume that there are two agents: *agent A* and *agent B*, each judging that *object x* is evaluatively good. On the Kantian picture, it might be said that *object x* provides a determinate instance of happiness. It follows that both will vie for *object x*. However, *object x* is a scarce good. To complicate matters further, we can assume that both agents make the evaluative judgement that *x* is good, but that each agent makes this judgement in order to satisfy diverging intentions. In other words, the pro-attitude (or the evaluative judgement) determining that *object x* is to be brought about is held fixed across the domain of agents. But, the attitudes determining intention and the noncognitive attitudes which determine the agents' expected feelings of pleasure or satisfaction will still vary. How should this case be adjudicated?

The mere fact of coincidence in the agents' evaluative judgements says nothing about the deservingness or otherwise of each agent. The coincidence says nothing about the best intended purpose, or even the description under which *object x* would be most valuable given full information. When the agreement of multiple judgements coincide, but this coincidence is dependent upon diverging intentions, there cannot be intersubjectively shared reasons for how to resolve the situation. Admittedly, this presumes that agents do not have qualitatively commensurable desire sets. But this is simply an assumption of Kantian philosophy. The stochastic nature of inclination makes it highly improbable that two agents would ever have qualitatively equivalent psychological responses, even when their evaluative judgements are held fixed.

My framing of the issue here largely relies on the interpretation of Kant which claims that incentives of inclination are not features of the world but psychological causes that suggest pleasurable experience. As I noted in the last chapter, the SCC theory does not share this interpretation. However, despite the different framing, I presume that the SCC theory would make the same claims in the case currently at issue. The framing for the SCC theorist would rely on the notion that, in the *private use of reason*, any agent is unable to know the desire set of another agent. I explain the conception of reason in its private use below.

Kant illustrates this kind of conflicting interest case in the *Second Critique*, where he writes in the voice of King Francis I to his brother the Emperor Charles V: "What my brother Charles would have (Milan), that I would also have" (V, 28).[40] Both brothers have judged the same object (Milan) to be choiceworthy, but they cannot occupy the region simultaneously. Specifically, this is meant to show that desires of inclination cannot be sufficient to act as universal law, even when these desires are in general agreement across the domain of rational beings. The agreement of inclination is merely contingent and can still result in competition, or worse, outright conflict. The agreement of reason, on the other hand, suggests a necessary purpose that all would attempt to bring about together insofar as they have practical reason.

If the SCC theorist wants to provide a universal, rational purpose or representative end of agency, then this role cannot be played by intersubjective agreement alone. There must be some feature which makes a reason intelligible not just as explanation, but as practical justification. In other words, there must be some feature which renders a reason truly *public* rather than *private*. In Kantian philosophy, a *public reason* is one that can be universalized, while a *private reason* is one which solely refers to the agent's own purposes. For the SCC theory, the feature, shared by every agent, which is capable of treating reasons as having universal normative force is the rational nature of agency itself. The act of making evaluative judgements, the SCC theorist argues, converges on the same objects because rational agents are all the same type of being. This provides an explanation of converging judgement which does not rely on mere coincidence and contingence.

2.3 Agency as a Final Good

In order to adjudicate conflict of interest cases, the SCC theorist appeals to the underlying structure that governs the reasoning of certain types of beings. Rational agents, for example, self-consciously endorse certain desires that determine the representation of something to be brought about. Evaluative judgements correspond with evaluative goods, the things judged to be choiceworthy. In this section, I will begin

[40] See also Korsgaard's comments about conflicting systems of private reasons at (2009, pp. 191–192), where she supposes that two agents vying for one object might conclude that they are in conflict.

to answer the questions raised before: what is the function of rationality? And, what is the characteristic function of agency?

Suppose, for example, that a self-conscious agent begins to probe his first-order desire endorsements, asking why it is that he endorses certain judgements about choiceworthy objects in the first place. In this case, a space for further reflection opens up for the agent's investigation. Self-reflective reason subjects itself to scrutiny. In response to the tendency of reason to continue investigating its own assumptions until a satisfactory answer is found, Korsgaard posits a further type of judgement that grounds the evaluative judgements of each agent. Without an adequate ground for evaluative judgements, a regress looms. Without an adequate ground, the agent might ask why he should engage in the activity of rational choice in the first place. The SCC theorist suggests that the only possible judgement which could terminate the regress is one which judges the agent's own rational nature to be good. Korsgaard describes these judgements as constitutive of *the final good* (2013). This is a type of judgement whereby the agent stands in a self-conscious relation to his own first-order, evaluative goods and judges this relation to be good as well.

Specifically, on the distinction between evaluative and final goods, Korgaard writes, "My claim is that the final good is grounded in a relation in which conscious animals stand to their own evaluative goodness: they are motivated to monitor and attend to it, and, in that sense, to make it their end" (2013, pp. 20). To illustrate the point that I take Korsgaard to be making, I will suggest an example that is perhaps closer to the orthodox reading of Kant than the SCC theory, in that I emphasize an epistemic capacity that does not depend on the capacity for agency. I will not claim that the example exactly illustrates Kant's position, though.[41] Specifically, my example will focus on the notion of *epistemic virtue*. The structure of the example will generalize to practical reasoning cases. I will hold off saying more about how the example generalizes until after I have presented it.

Suppose that, not unlike humanity, there is a type of rational being with the capacity to make certain theoretical judgements. For example, this being would be able

[41] Specifically, I want to avoid debates about the role of logic, theoretical judgments, and the understanding within Kant's philosophy. Hence, I am not claiming this presentation exactly accords with Kant.

to follow the inferential rule modus ponens, and judge *that q* validly follows from the two premises *if p then q* and *p*. Suppose further that this rational being makes a judgement about this very capacity for inferential rule-following. The active exercise of this capacity is judged to provide *epistemic reason*. The reason follows, not least, because the rule of inference ensures that, when the variables are replaced with sound premises, the rational being will represent the world accurately. Now, every time this rational being follows a rule of inference, the rational being will judge the conclusion to be *epistemically justified*. Perhaps the rational being would gain intellectual pleasure from following the rule, but this is not a necessary assumption. All we need to assume is that the rational being has a capacity which justifies following certain rules.

The above was merely the first step in the movement from *epistemic justification* to *epistemic virtue*. In summary: the rational being has some capacity and its exercise provides reason. The capacity provides the rational being with an ideal function. Since the capacity's function is characteristic of who or what the rational being is, and since the rational being can apply the capacity's rules well or poorly, then the rational being has reason to perform the function well because this would express to a greater extent the rational being's nature. Let us assume for the time being that the capacity for inferential rule following is properly constitutive for our hypothetical rational being. The capacity is constitutive of this rational being's cognitive makeup such that he can only aim at drawing conclusions from rules of inference. Of course, in this scenario, the rational being might fail to draw a valid conclusion, but he cannot fail to aim at drawing a valid conclusion.[42] Assume that our hypothetical rational being receives a blow to the head, or whichever body part houses his ability to perform rational functions. For the time being, the rational being would fail at providing justified conclusions that follow from his constitutive capacity. The possibility of the capacity's failure allows the rational being to experience the exercise of the capacity, not as determinate fact, but as *epistemic reason* to perform the function in question and to perform it well.

The movement from *epistemic reason* to *epistemic virtue* involves the self-conscious apprehension of a capacity's activity. The rational being can fall into the merely unconscious activity of rule-following, in which case the rational being might

[42] I say more about the failure of a constitutive capacity's exercise in the next chapter.

mistakenly apply modus ponens in a scenario which actually calls for modus tollens. In that case, the rational being runs the risk of providing an epistemic reason that is not truly justificatory. But, as a self-conscious rational being, he has the higher-order capacity to apprehend his own rule-following behavior and the ability to judge which exercise of the capacity would provide the best justification. This self-conscious apprehension of one's own capacity is the *relation of self to self* that Korsgaard is referencing in the above quotation. On the first-order level, the rational being has certain capacities, the exercise of which are judged to provide reason. But, as self-conscious, the rational being has the further, higher-order capacity to judge that the continued exercise of these capacities is itself a mark of *epistemic virtue*, and that the exercise of these capacities should be sustained.

Now, this brings us to an inquiry concerning how the above example generalizes to the practical sphere, and how the SCC theory frames its conception of the *final good* for every rational being. Specifically, the SCC theory focuses on one capacity, that of *rational agency*, which encompasses the ability to set ends for oneself, reason about how best to achieve these ends, execute actions, and provide intelligible reason for these actions. The first-order function of agency is prudential efficiency, such that the agent judges his own ends to be evaluatively good in some non-moral sense. For example, the agent might have a desire for some cake, endorse this desire, and then provide intelligible reason for the action of attaining said cake. If there are no defeating considerations, the attainment of cake would be *practically justified*. But, similar to our hypothetical rational being's ability to attend to his own epistemic functioning, the agent's higher-order function allows for the ability to step back and consider whether the desire for cake could truly act as the best justification. Does the desire for cake truly sustain the continued functioning of the agent?

There are several questions here about prudence. For example, how much cake can the agent eat without endangering his health? I am bracketing these questions because I am merely attempting to illustrate the self-conscious relation of self to self that Korsgaard sets forth.[43] The self-conscious relation by which the agent inspects his

[43] But see (IV, 399) where Kant sets forth his famous gout sufferer example to show that we have indirect duties to secure contentment so that we are not tempted to transgress direct duties. See (Timmermann,

own first-order evaluative judgements leads to the higher-order judgement that the ability to set first-order ends is itself evaluatively good. The judgement that the capacity for agency itself is good entails that agency itself is the *final good*. Agency, as a capacity, is the condition of having first-order desires at all. Therefore, this view which treats the capacity for agency to be philosophically primitive must assume that agency itself is the only true candidate for final goodness. The recognition of this capacity within oneself creates a different set of considerations: can a first-order desire be morally justified? Can the desire to do some act really be treated as the best reason that one has? Would other agents, in relevantly similar scenarios, also conclude that the agent has justificatory reason? Similar to the case of *epistemic virtue* above, the judgement that agency is the final good allows for the possibility of *practical virtue*.

The higher order identity of rational agency is the condition for the first-order activity of valuing. In other words, valuing is a characteristic activity of rational agency, and in every act of valuation, an agent expresses an affirmation of his own rational agency. This is why the constitutivist argues that the higher-order function of agency is *inescapable*. Every agent must endorse his own rational agency because this is the representative function of the type of being that he is. In this context, the final good is the agent's continued existence and ability to exercise its own characteristic activity as a valuer.

When situated communally, agents can choose to cooperate for the mutual recognition of the final good and virtuous function of each, or this cooperative venture could break down into vicious competition. The virtuous function of a single agent supports the final good, and happiness in proportion to this virtuous function is the *highest good* (Korsgaard,1996a, pp.118-119, 241; 2009, pp.87–88).[44] By describing the necessary evaluative judgements made by each agent, I have attempted to render explicit the conceptual connection between these evaluative goods and the final good.

The communal relation between agents creates a dilemma because each agent will pursue his own final good, but the most efficient way for each to pursue his own final

2022, pp.22) for a reading of this passage that suggests prudential reasoning is always uncertain because happiness is never guaranteed.

[44] For Kant's highest good, see (V, 110-111).

good seems to be that each pursues the final good of all. This is the constructivist solution to problem solving, whereby a solution to a practical problem—in this case the problem of mutual benefit—marks out the set of reciprocal responses that could possibly be provided (Korsgaard, 2008, pp.302–326; 2009, pp.188–196).[45] An agent's relation with certain evaluative goods imparts an identity onto the agent as *that type of agent who values these things*. In turn, this identity and the relations which constitute it are grounded in the inescapable identity of agency. The final good of a self-conscious agent is the agent's own characteristic functioning. The final good of each agent is then further conditioned by the relation to other beings who share this same characteristic functioning.

3.1 The Conditions of Choice

While I have outlined the general benefits of relationism and its application within the SCC framework above, I will turn, in this section, to a further description of the theory's assumptions. In particular, this section will attempt to connect some of the foregoing theoretical features from the previous sections to certain key passages in Kant. In other words, this section will attempt to show how the SCC theory can claim the Kantian moniker. The SCC theory has gone through significant changes on this point, and as such, I will begin by discussing the theory's development from the essays collected in *Creating the Kingdom of Ends* (Korsgaard, 1996a) to Korsgaard's more recent work (2008, 2009, 2013, 2018).

To see how the SCC theory first utilized the idea of Kantian relationism, one can turn to Korsgaard's interpretation of the passage from *Groundwork* II in which Kant introduces the Formula of Humanity (FH). Korsgaard sets forth her interpretation in the essays "Kant's Formula of Humanity" and "Aristotle and Kant on the Source of Value" (Korsgaard, 1996a, pp. 106-132, pp. 225–248), where she details "Kant's regress argument" and the notion of relationism. For the time being, I will focus on the regress argument. The argument concerns the passage from *Groundwork* II, Akademie edition

[45] For Korsgaard's (2008, pp.322) interpretation of constructivism, see the following: "[Normative concepts] are the names of the solutions of problems, problems to which we give names to mark them out as objects for practical thought." For other interpretations, see (O'Neill, 1989, 2006; Rawls, 1980; Reath, 2022; Schafer 2015a, 2015b; Street, 2010). For criticism, see (Enoch, 2009).

volume IV, pages 428-429, in which Kant considers several candidates as the source of unconditional value. The candidates run as follows: objects of inclination, inclination itself, non-rational beings, and rational beings.[46] The orthodox reading of the passage interprets Kant as arguing by elimination. In other words, the orthodox reading suggests that Kant lists each candidate and considers their benefits before eliminating each as a viable candidate before alighting on the universal end of rational beings (Timmermann, 2006; 2007, pp.94–98).[47] By contrast, Korsgaard reads this passage as a "regress" argument, by which the conditions for the possibility of something are explored (1996a, pp.117, 120–124). Since Kant's inquiry in the FH passage concerns the possibility of an end that could be represented as universal and necessary, the supposed regress argument explores the conditions of a universal account of valued ends.

The essence of the regress argument can be captured by expressing a single question that governs the characteristic inquiry of reason itself: what unconditioned rational principle does rationality provide to itself? To satisfy this inquiry, reason traces the conditions of its target until it reaches the unconditioned, which acts as a unified principle governing the systematic unity of the target cognition and other similar cognitions into a rational system, or *Wissenschaft*. In theoretical philosophy, the cognitions are immanent objects. In practical philosophy, the cognitions are such things as practical freedom and moral goodness.

There are two types of principles or ideals that can play the unifying role. There are *constitutive ideals* and merely *regulative ideals*. A constitutive use of reason is, as the name suggests, a principle of reason that constitutes the object of cognition (A179/B221).[48] The regulative use of reason is a principle which ensures that reason does not overstep its own boundaries (A180/B223). By reading the relevant *Groundwork* II passage (IV, 428-429) as a regress argument, Korsgaard is interpreting Kant as suggesting that the rational choice of ends is governed by the categorical

[46] Ralf Bader (2023) has suggested that the elimination argument does not take both "objects of inclination" and "inclination" to be relevant candidates. I am bracketing this concern because it does not change the present argument.

[47] I will not, here, enter into the dispute about how best to read this passage.

[48] In the context of theoretical philosophy, Kant writes that constitutive uses of reason "seek to bring the *existence* of appearances under rules *a priori*" (A179/B221), while regulative uses of reason are rules "according to which a unity of experience may arise from perception" (A180/B223).

imperative. In other words, this reading suggests that moral and non-moral value are constituted by a constitutive principle of practical cognition. So, for Korsgaard, Kant's question here is, what are the conditions such that the phenomena of rational choice and the activity of valuing are possible? The regress strategy then moves from the objects of inclination, to inclination itself, to non-rational beings, and finally the regress halts at the unconditioned nature of rationality itself.

The rational nature of humanity must bring its own practical cognitions under the a priori principle which constitutes intelligibility. This formal principle is constitutive of the unconditioned nature of rationality. When an agent cognizes something as a practical reason, he is thinking some material content through the constitutive principle that allow this content to play the role of an intelligible reason. This view presents every exercise of practical reason—from moral cognition to mere prudence—as aiming at the constitutive function of rational agency, which is determined by the principle of practical reason. The non-moral ends of agents are conditioned and suitably constrained by the reasons all agents can share. The aforementioned intelligible system of practical cognitions is how the SCC attempts to construct the solution of mutual benefit. Specifically, this is how Korsgaard interprets the Kingdom of Ends and the highest good for each (Korsgaard, 1996a, pp. 119).

The constitutive function of reason in its practical use is to allow for the cognition and universal intelligibility of practical reasons. Therefore, reason brings with it certain formal conditions applicable to the material content of experience. In other words, rational agents are beholden to the self-conscious structure of rational inquiry, such that some considerations can be thinkable as reasons which are shareable. This is why rationality is termed the "unconditioned condition." On the one hand, reason provides the conditions for cognition. On the other hand, reason itself is not known as a directly cognizable substance, but known through its own activity. Thus, reason is not itself constrained by conditions that come from outside itself but only constrained by the conditions which it itself imposes.

From the above, it is clear to see that the SCC theory has a Kantian account of intelligible reasons. But, does the SCC theory provide a Kantian account of the act of valuing? How is it that material content is thought through the constitutive rules of

reason such that singularly valued objects might serve as intelligible reasons to all agents? Specifically, Korsgaard takes the activity of valuing to be the rational choice of some object by an agent. This is the basis for the SCC theory's account of both moral and non-moral value. One can see this is so by turning to Korsgaard's discussion of *good* as a rational concept (1996a, 115). The reading of Kant that Korsgaard proposes relies on the *Second Critique* passage in which Kant discusses "the old formula of the schools": we desire nothing except under the guise of the good (V, 59-60). For Korsgaard, this passage suggests that, "Insofar as we are rational agents we will choose what is good – or take what we choose to be chosen as good" (1996a, 115). However, the worry here is that this reading obscures the very distinction that Kant is attempting to belabor in this passage.

"Choosing what is good" and "taking what we choose to be chosen as good" are two very different things. The latter—taking what we choose as something to be chosen—is merely in line with what Kant terms *the agreeable* (or *das Angenehme*). As sensibly affected beings, we all desire what we believe will make us happy. This does concern value, but only non-moral value. By contrast, the former construction—choosing what is good—is about desiring what can be judged antecedently to be *good* (or *das Gute*). In the case of the agreeable, the desire determines what is choiceworthy. In the case of the good, the judgement determines what could possibly be desired as good.[49] By equating the act of judgement with the activity of desiring something, Korsgaard ascribes a guise of the good reading to Kant that suggests the agreeable is conditioned by the good.[50] In other words, Korsgaard's SCC theory begins from the suggestion that agents rationally choose certain objects. The agent, then, moves from a desire to the judgement that the object desired can fit coherently within a system of intelligible reasons. For Korsgaard, the judgement that some desire could be shared universally by all agents within similar conditions chiefly means that the judgement is taken to be a reason intelligible to the faculty of desire of everyone (1996a, pp. 116). Hence, in the case of the SCC account of moral goodness, the circularity worry, raised

[49] See Timmermann (2022, pp.126–128) on a reading of the *Second Critique* passage which does not rely on a "guise of the good" interpretation.

[50] In a sense, this is true, of course. The good conditions what a virtuous person could justifiably pursue. But the good cannot fully condition what a virtuous person experiences as agreeable.

in section 2.2 above, is rendered less pressing by the *public use of reason*. On this account, if an action is judged to be good, then the agent should be capable of justifying this action to others (2009, pp.188–197).

As Kantian accounts of moral value go, there is one main problem with the above interpretation. The SCC theory gets the ordering of value judgements wrong, at least within the relevant Kantian context. Desire does not precede judgement in the case of moral value. By putting desire first, Korsgaard is able to model judgements of goodness on the mere act of taking some feature to be choiceworthy, or as she puts it, *taking what we choose as to be chosen*. The SCC theory's framing of this issue allows for the circularity worry to take hold. Kant, of course, does say that inclination "always has the first word" (V, 146-147). However, if we follow the classical interpretation and keep the use of empirical reason separate from the use of practical reason, then it becomes clear that Kant is not suggesting that desires or incentives of inclination are the rationally chosen contents of practical judgements.[51] It is true, though, that inclination has the first word in the sense that it gives the first suggestions of possible courses of action, which can then be judged permissible or impermissible through the use of pure practical reason. But, in the case of moral goodness, the circularity worry should not even get off the ground.

One could argue that Kant does allow for a desire that is specific to rationality and the capacity for the representation of practical ends. This objection would certainly be accurate. To see this, we can turn to Kant's introduction to *The Metaphysics of Morals*, where he distinguishes between desire in the general sense, desire in the narrow sense, and a determination of the faculty of desire through an interest in principles of pure reason (VI, 211-213). Kant writes that, "The *faculty of desire* is the faculty to be, by means of one's representations, the cause of the objects of these representations" (VI, 211). Thus, the faculty, or capacity, of desire as the ability to be a cause of represented objects is *desire in the general sense*. Kant a little later provides a definition of *desire in the narrow sense*, or desires of inclination: it is the "determination

[51] Complicating this matter further is the context of the passage (V, 146-147), where Kant proposes a thought experiment in which a rational being has luminous access to all questions of speculative philosophy, such as the existence of God, etc. The relevant point of the passage is to show how the moral law, in this scenario, would not be followed for its own sake. I will not deal with this more here.

of the faculty of desire which is caused and therefore necessarily *preceded* by such pleasure" (VI, 212). Finally, in setting forth the interest in principles of pure reason, Kant is clear that the interest of inclination cannot be substituted for an interest of pure reason. As Kant says, an intellectual pleasure, or habitual desire formed for the principle of pure reason, "would not be the cause but rather the effect of this pure interest of reason, and we could call it *sense-free inclination*" (VI, 213).

When discussing the rational choice of ends, Korsgaard blurs the distinction between the desire of rationality and desires of inclination. For the SCC theory, judgements of an action's goodness, or candidacy as a reason, happen only once some incentive has been picked out as relevant. In this way, the desires of inclination suggest certain features of the world, and rationality passes a judgement on the status of these features as reasons. Therefore, this explains why, for Korsgaard, the judgement that something is valuable is the same as the judgement that something is to be chosen. The view must concede that there is no real distinction between the function of desires of inclination and desires of rationality.

Although, as I have shown, the SCC theory does not adequately distinguish between types of desire, the regress argument does make clear that Korsgaard distinguishes between *a reason* and *inclination* as such. Under this interpretation, the material content of both reasons and inclinations are the same, but as has been discussed in the previous chapter, a reason is what is endorsed or constructed by the agent. If we frame this in the language of values, reasons are formed from material content—certain features suggested by inclination as *incentives*—and form—the constitutive rules of practical reason through which some desire is thinkable as a practical reason. Thus, the rational choice of objects transforms them from mere inclinations to *defeasible goods*.

The various classifications of value judgements, both moral and non-moral, are unified under the shareable system of reasons. When an agent steps back to ask whether some object could really be considered worthy of choice, and thus valuable, the agent is asking whether all other agents would take this object to be choiceworthy given similar conditions. An object worthy of *moralized choice* is one which every agent would necessarily agree to choose, supposing that the agent grounds his choice in the correct

principle of judgement. An object worthy of *non-moral choice* is one which is permissible given the constraints imposed by the relevant moral considerations.

The constraints on what action or maxim could be morally chosen are determined by the principle of practical reason. Further, moral choice conditions the validity of evaluative judgement. An object can only be thought good, in a non-moral sense, if it conforms to the conditions set out by the principle of practical reason because only then could it be thought as a reason. Of course, this depends on an assumption that the final good of humanity is determined by the functional nature of rational agency.

3.2 Conferring Value onto Oneself

In much of Korsgaard's early work, specifically the essays collected in *Creating the Kingdom of Ends* (1996a, esp. pp.106–132), it would appear as though Korsgaard takes a substantialist line. Value, Korsgaard argues, is conferred by human beings because human beings are the only beings who possess a rational will. This has more than a passing similarity to the substantialist view. Take, for example, someone who asks why the rational will is able to confer value onto the objects it chooses. What makes the rational will capable of value conferral? In response, the SCC theorist could respond by appealing to the regress argument: in search of something that could be treated as a universal end, rationality itself halts its search when it comes to the constraints set by humanity's own rational nature. Therefore, the will can confer value onto objects because the will itself is valuable, or at least, that is how the thought would go.

The exact time when Korsgaard changed her argument on value conferral is unclear, and the timing matters little for the current argument.[52] What does matter is the type of consideration that would have motivated a change in the SCC theory. One such consideration can be seen by turning to a prominent counterargument suggested by Rae Langton (2007). The fact that some subject confers a property onto an object does not mean that the subject instantiates the property in question. Examples abound,

[52] See (Korsgaard, 1998), "Motivation, Metaphysics, and the Value of the Self," where she states the change in her view explicitly. See also (Korsgaard, 2021) for a more recent discussion.

including Langton's apt statement that, "We have no more antecedent reason to expect the creators of goodness to be good, than to expect painters of blue to be blue, or the creators of babies to be babies" (2007, pp. 175–176). Equally, an electorate might confer political office onto a democratically elected candidate, but this does not mean that the electorate themselves occupy the political office in anything more than a metaphorical way.

In order to salvage the interpretation of Kant as a relationist, Korsgaard changed her assumption about the instantiation of value within humanity's rational nature, eventually moving to the idea that the rational will even confers value onto humanity. Although this method does get around the issue raised above, the new assumption raises further questions about how this type of self-valuing works in practice. What type of value could this be, such that an agent does not simply stop valuing itself and become less than an agent? What ensures that the rational agent inescapably values itself? To posit self-valuing as the description of some unperceivable explanatory phenomenon—such as the way that natural science observes the pressure exerted by a gas within a chamber and explains this by postulating the unobservable entity of a subatomic particle[53]—would fail to explain anything like the phenomenon of valuing. Valuing is an irreducibly normative activity and cannot act like other descriptive explanations in natural science.

For example, it would not be enough to experience the general fact that most people value themselves and then attempt to explain this by appealing to the necessary fact that humans rationally value themselves. The question that needs to be answered is this, why is it rational for humans to value themselves? Positing the activity of self-valuing as a brute fact does not explain why self-valuing is rational, and it cannot justify the activity of self-valuing. Brute explanation would only explain that there is a mechanism whereby value does take place. And simply positing a mechanism that causally leads to the phenomenon of value tells us nothing distinct from the fact that we have taken value as the target of our inquiry, something which, of course, we already

[53] See Sellars (1956, pp. 30) for the use of this analogy in showing how logical positivism fails to provide a genuine *theory* of linguistic usage. The subatomic particle analogy is meant to illustrate theoretical explanation, which logical positivism does not provide.

know. Instead, the question must be formulated in this way: if value is normative and choice-guiding, why is it necessary that rational beings choose themselves? What are the conditions such that rational beings are justified in valuing themselves?

I have already gone some way to answering these questions in sections 2.1, 2.2, and 2.3 above. There I argued that the representative function of agency allows for the rational choice of individual objects of desire. By expressing the affirmation of a first-order desire, an agent is also expressing the affirmation of this first-order desire's possibility, or his own identity as rational agent. Rational agency needs to act as the final good in order to stave off a vicious regress. To justify the normativity of this claim, what needs to be shown is that agency itself is a choiceworthy end that all agents must universally share.

The higher order identity of *rational agency* provides the possibility of first-order identities, such as mother, teacher, friend, and so on. Agents constitute these first-order self-conceptions by endorsing first-order desires. By desiring to *x*, an agent takes on the identity of an agent that desires to *x*. By desiring to dance professionally, the agent goes some way to becoming a professional dancer. But there must be something further than merely desiring something in order to make it the case that this desire constitutes an identity. Desiring to dance is not sufficient to make me a dancer. This begins to clarify what the condition of self-constitution must look like. The condition must reflect the agent's ability to achieve the things it desires. This is a corollary of what has already been discussed in the last chapter as the *practical efficacy view*.

From the preceding discussion of the higher-order constraints imposed by rationality, we can attempt to define the grounding condition that allows for the cognition of practical reasons. I am referring to the object of our inquiry as a *grounding condition* because, as *the condition* of the agent's first-order desires, the grounding condition must allow for the possibility of first-order desires. And as *a metaphysical ground*, the grounding condition must halt the regress that threatens the SCC theory's inquiry into the source of desire. The grounding condition is simply the "unconditioned condition" in a slightly different idiom. Notably, the idea of a metaphysical ground does not need to be so robustly ontological. Some philosophers—for example, Shamik Dasgupta (2017)—argue for a *deflationary ground* which can provide for much of the philosophical

explanatory work that other more robust theories of grounding attempt to provide. It is more than plausible that Kant is engaging in a similar method when he frames the limits of metaphysical inquiry as bounded by the structure of rational capacities.[54] The question is not, what real entity do these capacities correspond with? The question is, what are the rational principles which structure these capacities?

For the purpose of justifying such rational principles, constructivists appeal to the self-conception of the agent, the standpoint of some practical reasoner, or an identity that all agents must share. In Tamar Schapiro's (2021, pp. 21) terms, the grounding condition would be a "guiding conception," whereby a participant is standing in relation to some activity which she is undertaking and chooses to affirm this guiding conception because it is found to be worth engaging in.[55] Similarly, Sharon Street has argued that the distinctive aspect of constructivist methodologies is the practical standpoint, or "the standpoint of a being who judges, whether at a reflective or unreflective level, that some things call for, demand, or provide reasons for others" (2010, pp. 366). As Korsgaard writes in Sources 3.3.1 (1996b, pp. 100), "The reflective structure of the mind is a source of 'self-consciousness' because it forces us to have a *conception* of ourselves. As Kant argued, this is a fact about what it is *like* to be reflectively conscious and it does not prove the existence of a metaphysical self." Evident from the above quotations is the emphasis that constructivists place on the first-personal structure of deliberation.

By following the constructivist method, the SCC theorist would ask, what activity are agents engaging in such that they are valuing and providing reasons for action? The principle must itself be rational if it is to structure the activity of giving reasons. Importantly, the notion of rationality at issue cannot be the same as *substantial rationality*, the concept which is used when an agent responds correctly to reasons. If both uses of rationality at issue were conflated, then we would have a scenario in which rationality dictates a reason to do something, but equally, the action would be rational because there is reason to do it. The SCC theorist suggests that the higher-order capacity for rational agency can get around this problem for two reasons. First, a principle of rational agency provides a *normative explanation*, where this means that the

[54] Schafer (2018, 2019a, 2019b, 2020, 2023) interprets Kant in these terms.
[55] This is how Schapiro (2021) interprets Kant's method more generally.

explanation is not only descriptive but also action-guiding. For example, the principle dictates that an agent *determinately performs* a rational action if it is done in ideally rational conditions with full information. But, outside of ideal conditions, the principle also dictates that the agent *should* perform the rational action, and that there are better and worse modes of performance.

Second, a principle of rational agency is *justified*, in the sense that agents must affirm this principle as a universal account of first-personal choice and action. Specifically, when rational agency is framed as a capacity for efficient action, then it is supposed to play the role of a higher-order ground that all agents universally share, and this is meant to be clear from first-personal reflection. When I take part in an activity of valuing and providing reasons for my actions, what is it that I am doing? Presumably, I am affirming my identity as a rational agent. The principle's affirmation allows for a rational self-conception which justifies the agent's endeavors.

It is assumed that every agent does have first-order desires. One natural fact of human agents is that they need food and water, for example. So, what principle would be universally worthy of affirmation and capable of guiding the activity of first-order desire fulfillment? In response to this question, the SCC theorist would suggest that each rational being with first-order desires, and the ability to execute actions freely for the purpose of achieving these desires, must endorse a conception of himself as an autonomous agent. There are two features of first-personal reflection which render the principle clear. First, I have desires, and I cannot simply will that these desires be obliviated. And second, I must act, either to endorse these desires or to attempt to repress them. Either way, of course, I am exercising agency. Thus, from the first-person perspective, the universal normative principle must be a principle of agency *because* I do conceive myself as capable of acting on my desires and endorsing them as reasons. The principle which rationally structures the capacity for agency is expressed in every action and attempt at desire fulfillment.

For Korsgaard's framing of this exact point, one only needs to turn to *Sources* 3.2.2, where she writes:

> The freedom discovered in reflection is not a theoretical property which can also be seen by scientists considering the agent's deliberations third-personally and

> from outside. It is from within the deliberative perspective that we see our desires as providing suggestions which we may take or leave. You may say that this means that our freedom is not 'real' only if you have defined the 'real' as what can be defined by scientists looking at things third-personally and from outside. (1996b, pp.96)

In this way, the SCC theory takes the normative explanation and rational justification of rational agency to coincide in the first-personal conception of the agent as such. From the affirmation of oneself as a free rational agent capable of endorsing desires, the agent takes on certain *practical identities* which allow for desires to be constructed into reasons (Korsgaard, 1996b, pp.102). And, in a commonwealth of agents issuing reasons by virtue of these practical identities, rational agency, from its own internal grounds, demands that practical reason should be treated as *public* rather than *private* (1996b, pp.132-145). In the next chapter, I will turn to a criticism of the view set out here. Specifically, I will suggest that the notion of autonomous agency at issue here triggers the common objection that rational agency is an insufficient to ground categorical moral authority.

4. Conclusion

In this chapter, I have attempted to set forth the relational theory of value, how this is meant to correspond with Kant's view, and how the view deals with certain problems. Much of this chapter was expository and serves as a segue to the argument I the next chapter. I have said very little about Kantian substantialism, the main philosophical rival of Kantian relationism, but I have raised some of the worries that substantialists often raise. I ended with a discussion of final value and the rationalist ground of efficient agency. In the next chapter, I aim to show that the ground provided by the SCC theory cannot provide the categorical moral authority that Kantians often want.

Chapter Four

1. Introduction

I concluded the last chapter by looking to Korsgaard's notion of freedom from the *Sources of Normativity*. With that presentation concluded, I will turn, in this chapter, to a common counterargument. The worry is that the SCC theory's account of rational agency is insufficient to provide for the categorical moral authority needed for a Kantian theory. Before I detail the objection, though, I make explicit some of the SCC theory assumptions which allow the counterargument to work. In the last section of this chapter, I detail an interpretation of Kantian constitutivism that gets around the counterargument.

In section 2.1, I outline something that Michael Smith (1994) has called the "moral problem," and I provide a general outline of what the Kantian response should look like by turning to his second gallows man example. In section 2.2, I outline the SCC theory's attempt at meeting the moral problem. In section 3.1, I provide the insufficient authority objection. Finally, in section 3.2, I provide the alternative interpretation of Kantian constitutivism.

2.1 Kant's Gallows Man and the Moral Problem

Throughout this section I will refer to an issue that Michael Smith has referred to as the "moral problem" (1994).[56] In particular, this is a problem for *practical cognitivists*, those philosophers who argue that ethical statements can have a truth value and that a true ethical statement is sufficient to play the motivational role. If, for example, an ethical statement can have a truth value, then it must be a matter of *belief*. This is a fundamental assumption of ethical cognitivism: an agent has an attitude taken toward some ethical proposition which has a referent. If the agent declares, I believe *that murder is immoral*, the proposition has a truth value, and the cognitive attitude taken toward the proposition allows the agent to represent its truth.

[56] For Smith's Humean-rationalist constitutivist hybrid, see (Smith, 2007, 2013, 2018).

But, how can a cognitive attitude such as belief motivate an agent? Someone might declare, I believe *that there is a burglar in the corner*. However, belief does not motivate. Belief merely allows for the accurate representation of the world. If one truly believes *that there is a burglar in the corner*, it is not the mere belief that motivates one to do something about it. In that scenario, one would be motivated by some non-cognitive attitude such as fear, anger, hatred, or the overwhelming sense of propriety that one should call the police.

The focus on *practical reasoning* in ethics has been proposed as a way to analyze exactly how it is that agents are motivated to act for a reason. To analyze practical reasoning as such is to analyze the first-personal notion of what one should do or has reason to do. The easiest way to theorize about ethical motivation—how it is that an agent is motivated to act in accord with common ethical norms—is to subscribe to non-cognitivism: ethical statements have no truth value and motivate by virtue of the non-cognitive attitude expressed, such as fear, anger, or any other non-cognitive attitude. Non-cognitivism provides a straightforward metaethical explanation of motivation, but of course, this would jettison the hope of treating ethical statements as truth-apt.

For this reason, many view non-cognitivism, at least in the crude form in which I have presented it, as a philosophical nonstarter. There is a clear lesson that cognitivists must heed from the non-cognitivists, though. Mere belief is not sufficient to play the requisite motivational role when an agent is reasoning first-personally about what is to be done. Some pro-attitude such as desire must be a further necessary condition. However, some version of cognitivism must be accepted if we want to claim that ethical statements have truth value. Practical cognitivism, then, is an attempt to get around the "moral problem." It seeks to show how an ethical statement might be both *cognitive* and *practical*, how it might both refer and motivate.

In order to see how it is that moral cognition can play the motivational role within Kantian ethics, it is useful to turn to the gallows man example that Kant gives in the *Second Critique* (V, 30). The first gallows man thought experiment is a non-moral case illustrating the way that rational beings reevaluate preferences when faced with conflicts of interest. But, for our current purposes, the second gallows man example provides a

more interesting case. The second gallows man thought experiment proposes a situation wherein a prince asks his courtier to give false testimony against an honorable man. The courtier knows that the accused man has committed no crime, and that the prince is merely searching for an excuse to kill the accused. It is assumed that the prince wants his courtier to commit false testimony in order to have a spurious justification for corporal punishment. In addition, the prince threatens the courtier with death if he refuses to comply. In this scenario, the courtier is faced with a practical problem. There are two courses of action, and the courtier cannot refuse to act. One course of action—compliance with the prince's draconian diktats—is determined by the courtier's love of life. The other course of action—refusing to perjure an innocent man—is clearly the moral action, but this moral action promises death by hanging (V, 30).[57]

Kant argues that moral cognition can determine practical deliberation, even when we are faced with a scenario that threatens punishment in return for moral action. The clear and distinct way in which we can cognize the moral course of action proves, for Kant, that we have this capability independent of causal determination by nature. However, simply because we have a specific capability does not mean that we necessarily will attempt to exercise it. One has the capacity to act in accord with moral cognition, and one can refuse to do so. This is a question of moral motivation and justification. Why is it that we do have reason to act ethically? And why is the moral reason the all things considered best reason?

The question just raised again brings forth the problem of the normative, which I raised previously (Korsgaard, 1996b, pp. 92–93). It is the question of which desires or incentives really *should* act as a reason. Korsgaard's answer to the normative problem is a specific interpretation of the *rational will* as the capacity to reflect and act as a self-conscious agent. By connecting moral cognition directly to the capacity for agency, the SCC theorist provides an explanation and justification of moral rationality that hinges on the ability to reason within a set of self-imposed restraints. For this response to work, the SCC theorist must show that, given the identity of all as rational agents, we inescapably aim at action that is done from within the specified constraints.

[57] The practical postulates in the *Second Critique* suggest that we can hope for reward after death, but we would never have theoretical knowledge of this (V, 122-134).

2.2 Agency and Inescapability

Return to the example, given in the last chapter, of the capacity for inferential reasoning. This capacity allows a rational being to reach true conclusions from valid and sound premises through the cognitive activity of deductive inference. On the explanatory level, the capacity operates determinately: from the premises *all men are mortal* and *Socrates is a man*, the conclusion must follow *that Socrates is mortal*. In an ideal situation, given full information, and the allowance that one's rational capacities can proceed uninterrupted, we would simply be able to describe the capacity's activity as behavior that follows from the set of conditions within which it is operating.

Now, the perhaps unfortunate, but all too real fact, that plagues rational beings who have evolved from fleshy machinery is that these rational capacities can fail. Humans are not *fully* rational beings, and as such, the rational function of a capacity is likely to be interrupted by the world, which sets a limit on the extent to which a capacity can be exercised. Suppose that a rational being with the capacity for inferential reasoning sticks his finger in an active plug socket. For a short while, this being is dazed, his rational capacities are muddled, and he becomes temporarily incapacitated. Now, when the rational being is asked, perhaps by onlookers who want to know whether an ambulance should be called, whether—given the aforementioned major and minor premises—Socrates is mortal, this once rational being might spurt a string of garbled nonsense along with a lot of drool. In this scenario, we know how the capacity *should* be functioning. But, the capacity has failed in its characteristic activity. It is this very possibility of a capacity's failure to exercise its own characteristic function that *justifies normative statements* with respect to this very capacity.

We can easily see how the above point applies to the SCC theory's account of rational agency. For example, in an ideal scenario, a rational agent would act freely and not commit an act of self-harm even when directed to do so. But, take the case of Jones who has a mischievous friend that laughs at mild cases of others' pain. The mildly sadistic friend takes Jones's arm and, using it as a tool, punches Jones continually in the face with it. Now, since Jones is a rational agent with the capacity to be an efficient cause in the world, cases like this contradict Jones's constitutive principle of agency.

Not only has Jones ceded control over his own capacity, but he has allowed it to be used for the purpose of self-harm, too much of which will damage the very capacity which allows Jones to self-harm. From this, it follows that Jones *should* function as an efficient agent, where this means that he should wrestle back control over his arm from his mildly sadistic friend.

The SCC theorist interprets the categorical imperative as a universal higher-order attitude—call it a *rational desire*—which conditions the possibility of first-order desires. This higher-order attitude entails a *cognitive aspect*, such as the judgement that one believes some act to be necessary or permissible. But this attitude also entails a *conative aspect*, such as the desire to carry out the activity that has been judged necessary or permissible. Therefore, the SCC theory takes there to be a tight conceptual connection between belief and desire. It is assumed that this higher-order attitude, the categorical imperative, is a principle which conditions the possibility of being an agent at all, and thus, it constitutes the agent's first-order desire set. It allows for the possibility of the agent even having first-order desires.

In the principle's descriptive mode, it determines the activity of rational agency under ideal conditions, such that the agent judges some act to be necessary, represents this act as such, and then takes the means sufficient to bring about this representation. In the principle's normative mode, it justifies the use of deontic statements with regard to any one of the many scenarios in which the exercise of this principle could fail. From the foregoing assumptions, we can formalize the details of the SCC theory's specific form of practical cognitivism:

> **Agential Cognitivism**: As the condition of having first-order desires at all, there is a higher-order attitude with both cognitive and conative aspects. The cognitive aspect of the rational desire picks out certain features of the world as relevant. And the conative aspect justifies the action of bringing about or sustaining these features.

As a capacity which all agents must share insofar as they are to act on first-order desires, the cognitive aspect allows for the functional description of some governing principle that is to be followed. The conative aspect, on the other hand, allows for failure

in the principle's execution, and thus, the possibility of failure renders the principle normative.

If the fundamental principle of practical reason is a higher-order attitude, then someone might object that this contradicts the earlier presentation of the principle as a higher-order decision procedure. Even if we bracket the worry that the principle, as attitude, cannot play both a cognitive and conative role, the worry might still be raised that an attitude cannot function as a procedure. How is it possible that one principle is capable of playing both the role of an attitude and a formal procedure?

This counterargument raises a valid objection, but it is one that is easily disarmed by attending to the difference between the principle as descriptive explanation and the principle as normative justification. For an ideal agent who exists in conditions which allow for full information, the principle of practical reason is constitutive of what *would* be judged as good and then enacted. This would be the ideal agent's higher-order attitude; hence, we would describe the ideal agent's behavior as determined by the cognitive and conative aspects of this attitude.

By contrast, imperfect agents can fail in executing the action that this principle describes. Imperfect agents do not have full information. To paraphrase Robert Burns: the world often thwarts the best laid plans of mice and men. For an imperfect agent, the principle of practical reason is constitutive of what *should* be judged as good and then enacted. In the normative mode, the principle, as a procedure that operates from the limited information available, suggests the best possible solution. This goes some way to explaining why Korsgaard has variously referred to her theory as one of "procedural realism" (1996b, pp.36–37) and also as a theory opposed to metaethical realism (2008, pp.302–326; 2009, pp.64–65). The procedure is meant to be objective but not independent of the agent's own reasoning.

This provides Korsgaard with the justification to interpret the Formula of Universal Law (FUL) as a universally authoritative practical procedure (1996a, pp. 77–105).[58] Under this interpretation, the procedure produces different determinations of the good dependent upon the conditions from within which it is taken up. Each new scenario will result in the production of maxims with important differences. The relevant

[58] For a similar constructivist interpretation, see O'Neill (1989, pp. 86–89).

sense of universalization, then, is a requirement that a maxim would be accepted by others if the conditions for its willing were made fully intelligible. Universalization is not an attempt to make one's maxim apply to everyone. If a maxim does not apply to a scenario, then the maxim is irrelevant.

In addition to the procedural interpretation, the SCC theory adds a fallibilist proviso. Call this *provisional universalization*. The assumption is that there are certain background conditions for the universalization of maxims which render some maxims acceptable and others not. For example, the maxim of becoming a doctor for the reason that I am simply so inclined—independent of anyone else's need—cannot be universalized. This maxim requires the background condition that society has a need for doctors. Humans are cognitively limited, and thus, cannot know all the relevant conditions which might render a maxim non-universalizable. Therefore, a maxim can be put forth as provisionally universal with the intention of revising the maxim should invalidating information come to light.

With the procedural interpretation set forth, and the SCC theory's fallibilist commitments, it seems then that there can be no such thing as an unintelligible act because all actions must necessarily aim at the principle of practical reason. This is what constitutivist theorists term *inescapability*. In essence, this assumption suggests that an *explanatory and justificatory reason* can be given for every action, even actions that fail to bring about the represented good and even actions that are not all things considered best. It is inescapable that one does operate within the normative domain,[59] but of course, simply because *a reason* can be given to justify and explain one's choice of action does not mean that the action taken was the best reason one had. Cognitive limitations sometimes preclude the judgement that would have been made in ideal conditions.

This is mainly why the SCC theory has been criticized as bringing the Kantian notion of evil too close to the notion of mere unthinkingness. If I act badly, the SCC theory explains this by suggesting that I unthinkingly willed a bad maxim. The higher-order attitude—as the rational faculty of desire discussed in a previous chapter—might

[59] Alternatively, pragmatists sometimes refer to this as "the space of reasons," so coined by the philosopher Wilfrid Sellars (1956/1997).

suggest which incentive can act as the best reason. The unthinkingness worry is that the SCC theory renders a reason too cognitive.

In what follows, I will bracket the unthinkingness concern, in part, because I have already addressed it in Chapter Two. And, in part, because I argue that the insufficient authority criticism is ultimately more dangerous to the constitutivist program. Why is the unthinkingness criticism less worrisome for the constitutivist? In effect, the SCC theorist will simply say that the unthinkingness criticism begs the question against the constitutivist metaphysics. If the assumption is that moral action only comes from sufficient reflection, then this is the account of moral action we must contend with. A critic would either need to say that reflective endorsement is not a requirement of rational agency, in which case this is really the insufficient authority objection by another name; or, a critic would need to argue that reflective endorsement is a bad model, in which case the SCC theorist would simply disagree. Either way, the unthinkingness criticism does little to destabilize the SCC theory on its own.

Critics often illustrate the insufficient authority objection by attacking the assumption that rational agency is inescapable. If rational agency is not inescapable, and rational beings can willfully choose to aim at some other end than rational agency, then it is simply not constitutive of one's identity. If ethical normativity is escapable, then the authority of morality has the same grip on rational agents as norms of custom and propriety, which is to say that ethics has only a contingent and general grip on rational beings.

3.1 The Shmoral Law Within

Before turning to the counterargument, I want to make a brief note about one framing of the counterargument which ultimately fails to target Korsgaard's theory adequately. This deficient framing suggests that Korsgaard is arguing with a skeptic that completely lacks moral concepts. In Sources, Korsgaard writes:

> The real threat of moral scepticism lies here. A moral sceptic is not someone who thinks that there are no such things as moral concepts, or that our use of moral concepts cannot be explained […] The moral sceptic is someone who thinks that

> the explanation of moral concepts will be one that does not support the claims that morality makes on us. (1996b, pp.13)

What, in effect, would it even mean to argue with someone who does not have the concept of a moral reason? If some agent exists outside of the moral domain, there is no way of showing, through discourse and argument, that this agent should take up the moral standpoint. There is no squeezing morality out of a standpoint which takes morality to be unintelligible.

A framing which suggests that Korsgaard is attempting to reason with this kind of skeptic is not the best way to frame the argument. In actuality, the SCC theory is presented as an inquiry into the conditions which allow for reasons responsiveness. Its method seeks to explain and justify how the cognition of moral reasons is even a possibility. The reasons themselves are *constructed*, but the conditions which allow for the construction of reasons are meant to be *constitutive*. A counterargument that takes as its target a form of constructivism that attempts to construct the constitutive conditions is simply not arguing against Korsgaard. Chapter Three explored how these constitutive conditions arise from the agent's relation with the world. Notably, the argument given there did not refer to a process whereby reasons are given for the construction of rational agency and the norms that govern it.

A better framing of the counterargument suggests that Korsgaard is attempting to argue with a skeptic who does not see why his own moral concepts hold authority. The skeptic, in this case, does not see what justification he has for acting morally. Suppose a person was raised in the confines of a repressive cult that proclaimed its own ethical system. One day, this person leaves the cult and finds that the guiding principles he was taught were all a sham. This person still has these principles, and perhaps they enter his conscious mind from time to time, but he no longer thinks they hold authority over him. Korsgaard's moral skeptic is someone like the person who has escaped from the cult. He recognizes moral concepts, and perhaps when he feels sufficiently motivated, he even acts on these moral concepts. But, he still contends that morality is no more authoritative than the cult's principles. This version of moral skepticism is how the insufficient authority objection proceeds.

In framing the counterargument, I am drawing from several prominent criticisms of the SCC theory, most notably David Enoch's "shmagency objection" (2006).[60] Largely, the shmagency objection is only original in its presentation. I will suggest that many of the prominent counterarguments are making the same point, only in a less stylistically attractive way. The main criticism, shared by the shmagency objection and other common objections, is that the SCC theory is simply not sufficient to purchase a universal account of ethical normativity, and it removes authority from ethical reasons. This is why these counterarguments often come from metaethical realists, many of whom suggest that the SCC account of practical reasoning can be saved by supplementing it with a metaethical realist account of normativity (FitzPatrick, 2005, pp.684–691).

There is something true to this counterargument, though I will ultimately argue that the criticism can be avoided without buying into metaethical realism. In this section, I aim to clarify why these arguments work against the SCC theory. Later, I will suggest that Kant does propose a version of metaethical constitutivism that gets around this problem, although his account falls prey to other worries that ultimately show the limits of constitutivism.

Before I show exactly how the counterargument works, I will turn to the distinction between constitutive and regulative standards. I should make one quick note before I proceed. The examples I use largely discuss the functions attributed to non-rational objects. Naturalist constitutivists often use examples that describe the natural function of some unconscious or conscious bodily capacity.[61] I am using similar examples because I have already addressed in Chapter Three how the SCC theory takes the agent to ascribe functions to non-rational objects.

In order to illustrate the difference between a constitutive and regulative standard, I will suggest two explanatory statements:

---

[60] But, see also (FitzPatrick, 2005, 2013). See (Enoch, 2009) for an attack on "global," thoroughgoing, or metaethical constructivism. For a presentation of these kinds of counterarguments as alienation from the "rational authority" of normativity, see (Samuel, 2022).

[61] See (Fix, 2020, 2021, 2023) for some examples, but also, note that Fix's point is that the function of these capacities only holds in relation to certain objects they interact with. Fix's point is *not* that these functions operate of their own accord.

A. You have a human nutritive capacity *because* you have the capacity to digest the organic materials necessary to sustain human life.

B. You need the recommended serving of vegetables to sustain your own functioning *because* you have a human nutritive capacity.

I will start by discussing statement A, which is an example of a *constitutive explanation*. In this case, the human nutritive capacity is a *real necessity*, as opposed to a merely logical necessity. Things could have been different, but under the conditions in which someone has a human nutritive capacity, it is constitutive of this capacity that it functions as it does. It is part of what the human nutritive capacity is that it does function by digesting the type of plant and animal matter that is non-toxic for humans.

By contrast, statement B is an example of a *regulative standard*. The human nutritive capacity provides energy to the systematic operations of the human body by digesting organic foodstuffs. If the capacity does not operate, then the body that the capacity nourishes will perish. The capacity, in this case, would no longer exist. This causal chain explains why it is that digestion of foodstuffs into nutrients and expellable waste is necessary for the continued functioning of the nutritive capacity. If there is a privation of nutrient dense foodstuffs for the capacity to digest, then the capacity will no longer exist.

These two types of explanation lend themselves to two different types of imperatival statements. The transition from explanatory to imperatival statement is justified because, as I have already shown, the account in question takes normativity to be a quintessential part of the metaphysics of capacities. However, these imperatival statements will differ in what they take to be the function of the modal verb.

When statement A, the constitutive explanation, is translated into an imperatival statement, it requires a modal verb that reflects the really necessary property of having a constitutive identity. For example, *humans must have a human nutritive capacity*. The capacity to digest food in the particular mode that humans do is part of what constitutes

the human animal.[62] If the constitutive standard does not apply, then we would say that this being lacks an essential property of humanity.

Statement B, on the other hand, is a merely regulative standard, and as such, it cannot take on a modal verb that reflects real necessity. Regulative standards only hold generally, not universally and necessarily. For example, it is only in general that a struck match fulfills its function of producing fire. The correct conditions, such as a state of affairs in which the match is dry, must be held constant for the match to produce fire, otherwise striking a wet match would not produce the intended effect.[63] If we translate statement B into an imperatival statement, we get the following: *given normal conditions, humans should eat the recommended serving of vegetables in order to sustain their own functioning*. This statement utilizes a deontic modal to reflect that the action is proper, where propriety in this case simply refers to the behavior that is characteristic of humans who live physically healthy lives. If one week a human animal does not follow the prescription expressed by the above imperatival statement, nothing much would happen. If a human animal failed to follow the prescription every day of his life, the human animal would probably not live a physically healthy life, where this means that he would have a life expectancy outside the standard deviation.

The imperatival statements for A and B make it clear that there is a difference between the normativity that comes from constitutive standards and regulative standards. Constitutive explanations provide the conditions that must hold for something to be the type of thing that it is. Regulative standards provide a description of the pathway that something should follow for it to meet some standard that is considered good, valuable, apt, or worthy.

Can something fail to live up to a constitutive standard? Yes, this is why capacities are framed in normative terms. A capacity must aim at the behavior governed by its characteristic function, but this is no guarantee that the capacity will in actuality achieve its function. For example, the human nutritive capacity must aim at functioning

[62] There are problems with regard to the constitutivist framing of human nature and the occurrence of disability. Nevertheless, a Kantian rationalist constitutivism handily works around these problems. Rational capacities do not correlate with cognitive ability.

[63] Korgaard (2009, pp.37–41, 114) discusses teleology, causality, and scientific explanation, although, of course, she does not use the frame of regulative standards.

well, but perhaps the human animal suffers from a temporary bout of indigestion. The capacity in this case has failed to *perform* its function, but it *has not failed to aim* at its function. This is why aiming at the capacity's function is *inescapable*. If one does not aim at the capacity's function, then one simply does not have the capacity, and it must not be constitutive.

Can something fail to live up to the regulative standard? The answer is almost trivially yes in this case. Return to the scenario of attempting to light a wet match. In this case, the relevant causal standard is that a match produces fire when struck. The causal standard requires that the correct conditions hold. If the conditions do not hold, then the causal standard fails. A wet match cannot produce fire by striking it. So, must a match aim at being a cause for fire? This does not hold necessarily. A regulative standard is merely a constraint imposed such that the intended effect is produced. It is almost definitionally true that a regulative standard is escapable. As such, a match will only aim at producing fire given that it has been placed in the correct regulative conditions which produce the intended effect. If the agent wants to produce fire, then he *should* see to it that his match is dry. This is a regulative standard, and it is not inescapable.

While it is plausible that something like the capacity for rational agency is constitutive of certain norms, other practical identities are clearly regulative standards. Agency is inescapable. An agent cannot lose his capacity for agency without ceasing to be what he fundamentally is. In fact, in Kant's terms, the speculative use of reason furnishes the regulative idea of the rational soul, which is a "merely dialectical concept" (A644/B672). Korsgaard's practical identities are not really what Kant has in mind when discussing the rational soul. Nevertheless, if we transpose Kant's distinction between the constitutive and regulative to the SCC theory, it seems that agency, as the rational ground, is justifiably considered the constitutive identity, while other practical identities are only regulative.

We can follow the argument further: agency is inescapable, but the same cannot be said for other kinds of practical identities. Suppose that Jones is attending a dinner party. There are certain norms that govern behavior considered proper for a dinner party. Largely, these reasons of propriety are prudential and are set because of social

convention. They are reasons that serve merely to lubricate the social interaction such that all participating members of the party feel comfortable. Assume that conviviality is a norm imparted onto someone who identifies as a dinner party guest. But, the choice of whether or not to aim at conviviality is entirely up to the agent. There is only *generally* a reason for conviviality shared among dinner guests because dinner guests typically and for the most part want others to have a good time. The norm in this case only holds authoritatively if, given relevant information about the scenario, 1.) the norm is known to apply, and 2.) the person with the practical identity is sufficiently accepting of the norm.

Jones might not heed the norm, for example, if he does not know that his host wants him to be convivial. Perhaps, only last year, the same host held a dinner party where the guests were told to be rude to one another. Equally, Jones might knowingly reject the social convention of conviviality. Perhaps he is simply an uncouth curmudgeon. But, even when Jones does not aim at the relevant norm, he is, though maybe unwanted, still a guest at the dinner party. Only after too much offense will the invite be rescinded. Therefore, this type of identity is similar to the regulative standard at issue in statement B. In *general*, the norm of conviviality holds authoritatively for dinner guests. But this is only given the right conditions, such as full knowledge and sufficient motivation, that will produce the effect in question.

With the relevant distinction set forth, I will now suggest that the prominent counterarguments against the SCC theory are really getting at the insufficiency of practical identities to ground anything like a universal and necessary system of ethical norms. In other words, the suggestion is that the SCC theory presents moral reasons as lacking categorical authority. Specifically, norms imparted by practical identities are escapable. Therefore, these norms are not constitutive of one's identity and are merely regulative norms justified by prudential reasons. This argument can be made clear by turning to one of Kant's examples from the *Second Critique*, which is functionally similar to the second gallows man example. The relevant example is that of an honest man asked by Henry VIII of England to perjure Anne Boleyn. This example comes at the *Second Critique*'s Doctrine of Method, Akademie edition volume V, pages 155-157, and I am choosing to focus on this example rather than the second gallows man case because the notion of identity is more clearly at play in the Henry VIII example.

Again, as before, we have a case where an honest man is asked by a member of the royal family to submit a false accusation against an innocent person. In this case, this dilemma takes place in Tudor England, where Henry VIII has asked for an accusation against Anne Boleyn so that he will have the apparent justificatory ground to divorce Anne and marry someone else. Henry VIII first provides blandishments, inducements, and gifts. When this fails to motivate, Henry VIII turns to making draconian threats of harm. Under a Kantian ethical analysis, we know what the conclusion of this story should be. It is immoral to provide false testimony against others. For one, false testimony is a lie, which violates a strict duty. Secondly, false testimony uses another person as a mere means so that the agent can escape the threatened punishments. Finally, false testimony indirectly contributes to a harm directed at another person.

The constitutive identity of efficient agency is capable of grounding something like strict and wide duties to the self. Perhaps, in the Henry VIII case, the agent has developed his talents such that he has become a valued member at court. We can even suppose that the agent saw the development of his talents as a duty. However, when attempting to ground a duty to the accused, efficient agency cannot provide the correct response. Remember that Henry VIII first offers the agent reward for performing the command. At this point in the story, there is as yet not a sufficient reason that counts in favor of accusing the innocent or telling the truth. So far, from the mere ground of efficient agency, both courses of action are under-motivated, meaning that reason can be given for either action.

Suppose that the agent denies Henry VIII's gifts. The next phase of the thought experiment sees Henry VIII threatening the agent. If the agent refuses again, he will face physical harm. Whereas the previous phase saw two undermotivated pathways open to the agent, this phase provides sufficient reason for one course of action. If duties are grounded in the identity of efficient agency, then the agent has an all things considered reason in favor of providing false testimony. At the very least, physical harm will partially incapacitate the agent for at least some time. In the worst case scenario—a death sentence—Henry VIII will incapacitate the agent fully. In either case, the duty to

protect one's own capacity for agency wins out. Care for others is defeated by the ground of efficient agency.

What practical identity could possibly provide the agent with reason to set aside his own identity as an efficient agent and perform an action despite the promise of harm? There are two possible candidates for the relevant practical identity. First, there are socio-political identities, or identities conferred onto someone simply by performing a function in some social group. Second, there are natural identities, which correspond to some naturalistic fact. In the Henry VIII case, the most relevant socio-political identity is the identity of courtier because it encompasses such factors as social status, political moment, and cultural mores. The most general of the natural identities is the identity of humanity. In effect, these two options allow for two distinct methods: the socio-political identity would attempt to generate duties from very specific cultural features, while the identity of humanity would attempt to generate duties from a general identity. The identity of courtier reflects the very specific socio-political scenario in which the thought experiment is taking place. The identity of humanity generalizes across temporal and spatial divisions.

The courtier identity generates the wrong conclusion. As a courtier, it would be expected that the agent swore some oath of allegiance to his king, Henry VIII. The agent might have some other socio-political identity that conflicts with his oath of fealty. Perhaps, like Thomas More, the courtier has another identity that precludes his acknowledging Henry VIII's claims. But, in this case, what determines the conclusion that one socio-political identity would win out over the other? It is entirely dependent upon the conditions in which the conflict takes place and the value that the agent imparts on one identity over another. If the agent believes that his identity as Henry VIII's courtier should determine his actions, then this would provide sufficient reason to give false testimony. If the agent believes that his identity as, say, a Catholic dissenter provides the determination to deny Henry VIII's claims, then this would provide sufficient reason to speak in favor of Anne Boleyn.

Notice that a focus on socio-political identity leads to a scenario in which the choice of action is contingent; it would be dependent upon ideology or the social discourse of an in-group. We can attain theoretical understanding of any course of

action that the agent might choose. But focusing on socio-political identities alone cannot provide categorical reason for moral action. At base, these types of identities are *regulative*, not constitutive, because they provide reason merely within the right social conditions. Therefore, these identities are escapable. The agent can choose not to aim at action from these identities.

Although socio-political identities cannot generate the correct conclusion in the Henry VIII case, perhaps this is an uncharitable criticism of the SCC theory. After all, Korsgaard never appeals to socio-political identity. The SCC theory is, unlike many of its constructivist counterparts, not a theory of political normativity alone, but a theory about the sources of normativity simpliciter. The most charitable criticism, then, would have to show how something like a normative conception of humanity itself is a merely regulative identity and not inescapable. This criticism would not have to attack the idea that there is some universally normative conception of human identity. The idea that human identity is merely regulative can also lay claim to universal normativity. Instead, the argument would only need to attack the idea that human identity is sufficient for ethical normativity. This is what I will claim: the identity of humanity, even when conjoined to the identity of efficient agency, is insufficient for ethical normativity.

Let us assume that the courtier does not appeal to his socio-political identity in attempting to respond to the demands of Henry VIII. Assume that, instead, the courtier, as a rational agent with the ability to self-consciously reflect, asks himself what is the *most human* course of action? Of course, if we want to provide a normative answer to this question, it is not enough to describe the behavior of a large sampling of humans in suitably similar scenarios. The *most human action*, in this case, must be an action determined by a normative conception of humanity.

In order to avoid contingency, this conception cannot be one that is rooted in some socio-political identity. For example, one would be unable, in this case, to appeal to the philosophically liberal conception of humanity. In order to appeal to this conception of humanity, one must first justify the underlying assumptions of humanity as a class of free, equal, and rational beings. To be sure, there may be reason for accepting the liberal conception of humanity, but before the question is prejudged, one must show why there is reason to think of humanity in those exact normative terms.

Before one has done the justificatory work sufficient to show why it is that humans do identify as free, equal, and rational, one will only be able to appeal to the contingent historical circumstances that have causally led to the liberal conception of humanity being a viable identity. Socio-political identities rooted in the historical development of some place are assuredly escapable identities.

Neither are we able to appeal to some normatively governed naturalistic conception of humanity alone. For example, maybe under an evolutionary-psychological theory, humans have evolved to be social creatures, whereby empathy for fellow humans—perhaps theoretically explainable by appeal to some neuro-chemical or other—determines that human action should ultimately be other-regarding. Still, this cannot guarantee the categorical nature of the ethical norm in question. The evolutionary identity is merely normative in the regulative sense. Normativity, in this case, arises from the *normal* behavior of the species. For example, in this case, humans would often act in accord with ethical norms because certain neuro-chemicals would normally, or typically, push them in the direction of that action. But again, probabilistic likelihood is far from universality or necessity. If the neuro-chemicals misfire and the human behaves without empathy, then, at most, all we would justifiably be able to say is that this cold, apathetic human *should act with empathy to be like the other humans with more typical behavior patterns*. Even more troubling are the limits of empathy. There does not seem to be a conception of empathy motivationally strong enough to generate the right conclusion in the Henry VIII case. The agent might feel sympathy for Anne Boleyn, who is surely about to face a tragic end, but sympathy cannot override the prudential reason an agent has to secure his own well-being in the face of duress.

The above is not the only viable way to frame the normatively governed naturalistic conception of humanity. For example, one could subscribe to a version of natural law theory, where the ethical norms that govern human behavior are simply taken to be the correct function of the natural world. Of course, this would be sufficient to purchase universality of ethical norms across the domain of all human agents. Perhaps it would purchase necessity as well. But, I will not, here, follow the thread of inquiry further into whether natural law is necessary in the relevant sense. The SCC theory cannot work with any conception of natural law which posits normativity as an

irreducible aspect of the world that ultimately points toward *the good*. This would contradict the constructivist assumptions of the SCC theory.[64]

By eliminating the possible candidate identities, I have been attempting to illustrate a dilemma that plagues the SCC theory's bid at generating duties to others. Efficient agency is surely sufficient to generate universal and necessary duties to the self, and even more, it is inescapable. But, efficient agency is insufficient to generate universal and necessary duties to others. The dilemma, then, is this: either we ground our ethics in the capacity for efficient agency, an inescapable identity which is surely necessary and universal, but insufficient to generate anything like ethical normativity. Or, on the other hand, we supplement the identity of efficient agency with specific practical identities, which generate duties owed to others. The normativity arising from practical identities is not inescapable, necessary, or universal. Famously, David Enoch (2006) stylized this dilemma by suggesting that a skeptic might come along and query whether someone must aim at duties generated by practical identities at all. Even if one notion of agency is inescapable, the notion of agency doing the most theoretical heavy lifting within the SCC account of normativity is escapable. Why identify with the ethically loaded version of agency, if one could just as well claim to be a "shmagent" instead?

There is a possible defense open to the SCC theorist, but ultimately, I will suggest that the defense fails as well. In a previous chapter, I argued that the SCC theory does not subscribe to the Mental Objects (MO) view of desires or internal reasons. From this, it follows that an agent cannot merely pick his own practical identities. In other words, there is no choosing whether to identify or not with the ethical norms that apply in virtue of one's own identity. The desires that determine one's identities hold, or they do not. For example, Jones is a human. He cannot change this fact. Therefore, an agent cannot reason himself out of the constraints set by a practical identity.

While it is certainly true that an identity cannot be simply discarded, this is not the correct notion of inescapability. Unlike the identity of efficient agency, practical identities

[64] I am only suggesting that a Kantian cannot accept a version of natural law which argues for some substantially realist and inherent law of the world. There is at least some sense in which Kant is working in the natural law tradition.

operate like the example of the dinner party guest. In that example, the guest is still a guest at the party, but he does not aim at the norms generated by the identity. The identity still holds, but the norms are not inescapable. Similarly, in the Henry VIII case, the agent might have the socio-political identity of courtier—an identity that cannot be simply rejected by desiring to be something else—but the agent does not need to aim at the norms that govern this identity. Like the identity of the dinner guest, the identity of the courtier is regulative: these identities generate norms only given certain conditions, and the norms only hold authority if they are accepted by the agent.

The above point about practical identities serving as regulative, not constitutive, can be seen in one key example from Enoch's (2006, pp.189) essay. Enoch asks the reader to assume that an agent is stuck playing chess and cannot quit. In this case, the identity of chess player is truly inescapable. However, the inescapable identity does not provide normative reason to act. Admittedly, this passage can be read in two ways.[65] First, it can be read as the counterargument against constructivism that I have already dismissed. It is true that moral normativity cannot come from someone who lacks moral concepts, but the SCC theory does not claim this anyway.

There is a second, more salient way to read the above passage that presents the argument as a sort of *reductio* which gets at the point I have been belaboring thus far. Under this second interpretation, the chess player example is a type of identity which one cannot reason oneself out of.[66] It is simply the case—perhaps like the knight in Ingmar Bergman's *The Seventh Seal*—that the agent is stuck playing chess. But even if the agent has the practical identity of chess player which cannot be discarded, the norms that govern proper behavior are escapable for two reasons. In the first instance, the correct conditions must hold, without which the norms would not arise. And, in the second instance, the norms must be accepted, without which the norm would have no authority. Now, in the chess case, the conditions have been held fixed, such that the agent *probably should* be playing like a typical chess player. But, the agent just does not feel motivated to follow the norm such that the agent is an *atypical chess player*. The

[65] I will not claim to know Enoch's exact intention in writing this passage.

[66] See (Tenenbaum, 2019) for the presentation of this part of Enoch's argument as "normative alienation." The agent is inescapably stuck with an identity he does not identify with. See (Samuel, 2022) for a Hegelian attempt to get around a similar problem he terms "social alienation."

key, of course, is that the identity still holds *even when* the agent does not aim at the norm. Thus, practical identities cannot be constitutive of ethical normativity.

The above counterargument raises several problems for any theory which attempts to ground normative ethics in the capacity for agency. As has been shown, efficient agency is insufficient for generating duties to others. When efficient agency is supplemented with duties generated by practical identities, the theory no longer presents a categorical account of ethical normativity. The reason the norms generated by practical identities are not categorical is that these identities are not constitutive but merely regulative. As such, practical identities generate escapable norms. If the theory cannot ground a universal and necessary account of ethical normativity, then it is a far cry from its Kantian foundations.

Due to the above problems that plague the SCC theory, metaethical realists often propose that those sympathetic to Kantianism and the SCC theory should reject *agential cognitivism* and support some form of *metaethical realism* instead. Although realism would certainly succeed in the attempt to provide a ground for universal and necessary ethical norms, I will suggest that it is not the only available option. To see that this is the case, one can turn to a different interpretation of Kant's constitutivism as an alternative to the SCC theory. This alternative Kantian constitutivism does not attempt to ground normativity in the identity or capacity for agency. In fact, I will suggest that identity only plays a secondary role on the alternative reading of Kant. Once this alternative view is outlined, I will present some possible recommendations that will help Kantian constitutivism move forward.

3.2 Kant's Constitutivism[67]

This section will proceed by proposing an alternative interpretation of Kantian constitutivism. In contrast to Korsgaard's agential approach to constitutivism given above, Kant takes the principle of practical reason to be constitutive of *pure practical reason*, not the self-conscious use of *empirical practical reason*. For this reason, I take Kant to be a rationalist constitutivist that does *not* emphasize rational agency. While I do

[67] I have found Henry Allison's entry on "Reason (*Vernunft*)" in the *Cambridge Kant Lexicon* extremely helpful in structuring my thoughts for this section.

not want to deny the possibility of Kantian theories that emphasize rational agency, I argue that agency is inadequate to act as a ground for categorical moral norms. This section will develop these ideas further and provide an alternative to agential cognitivism.

In the *Second Critique*, Kant writes, "The concept of good and evil must not be determined before the moral law (for which, as it would seem, this concept would have to be made the basis) but only (as was done here) after it and by means of it" (V, 63).[68] This is Kant's famous "paradox of method," whereby he seeks to define *the law* before *the good*. What does Kant mean, here, by attempting to define the good "by means of" the moral law? Korsgaard takes this to mean that humans, by virtue of the capacity for agency and rational choice, confer value onto objects. Under the SCC interpretation, choice precedes judgement; the *activity of choosing* precedes the *concept of to be chosen*. As has been shown above, this reading is controversial.

As an alternative interpretation, Kantian constitutivism proceeds, not from a faculty of rational choice, but by defining a capacity that serves as the transcendental condition of making valid practical judgements. In other words, Kant's metaethical constitutivism is preoccupied with pure reason in its practical use. To see this, one must attend to the distinction Kant makes in using the term reason. First, there is *reason in the broad sense*, which refers to the collection of all higher cognitive faculties or capacities, the functions of which are structured by synthetic a priori principles.

In the margins of Alexander Gottlieb Baumgarten's *Metaphysica* textbook which Kant used for philosophy lectures, Kant left a note that defines the concept of a rational capacity: "The inner principle of the possibility of action is the capacity [*Vermögen*]" (Reflection 3585, XVI, 73).[69] Many of the notes which Kant left in this book probably date back to 1764, well before his critical period (*Notes*, Guyer et. al., pp. 68). Despite the note possibly dating back to Kant's Leibnizian-Wollfian days, Kant continues to emphasize the philosophical importance of rational capacities into his critical period, and in fact, they still play an important role in Kant's philosophy after the development of

[68] This whole passage is italicized in Gregor's edition.
[69] This note is left out of the Cambridge edition of Kant's *Notes and Fragments*. The original Akademie edition has, "Das innere Princip der Möglichkeit des Handelns ist das Vermögen" (XVI, 73).

autonomy in 1785. In the critical period, these rational faculties or capacities (*Vermögen*) are the cognitive limits, by which theoretical certainty is bounded. Of these faculties, Kant writes in the *Second Critique*, "All human insight is at an end as soon as we have arrived at basic powers [*Grundkräften*] or basic faculties [*Grundvermögen*]; for there is nothing through which their possibility can be conceived" (V, 46-47).

Collected under the broad notion of reason, these cognitive faculties can be further defined through a tripartite functional distinction. These faculties are defined by the rational principles that determine what it is they are potentials to do. In other words, the capacities are defined by virtue of the characteristic activities which they perform. The first of the cognitive faculties is the *power of judgement*. The second is the *faculty of understanding*, which can also be described as the ability to cognize the material of sensibility under a series of determinate rules (A126). Finally, the third cognitive faculty is *reason in the narrow sense*, alternatively characterized as the *faculty of principles* (A299/B356), whereby thought is collected under the "highest unity of thinking" (A298-9/B355). If, as Kant argues in the Introduction to the *Groundwork* (IV, 388), the task of metaphysics is to supply principles for the distinct areas of philosophy, then pure philosophy must investigate the rational principles which structure the characteristic activity of these capacities in order to set out the justified limits of their use.[70] Part of the *First Critque*'s overarching project is to show that the only justified use of the understanding is its immanent use. The *Groundwork*, by contrast, seeks to determine a principle sufficient for pure moral philosophy or a metaphysics of morals.

As the faculty of principles, reason has both a *real* and a *logical* use for theoretical inquiry. Although distinct uses of reason, the real and logical uses must both conform to a single principle of reason as such. The principle can be framed in this way: reason must "find the unconditioned for conditioned cognitions of the understanding, with which its unity will be completed" (A307/B364). In Kant's theoretical philosophy, the question of whether reason has a *real use* leads to an inquiry into whether reason is justified in its cognition of certain "transcendental ideas," which are the totality of

[70] See, for example, Kant's remarks on the same page (IV, 388) concerning the "barbarous state" of jack-of-all-trades professions. Before mixing the rational and empirical, one must first set forth the justified remit of rationality for the science in question.

conditions for a given conditioned thing (A321-322/B377-379). The transcendental ideas are threefold: the soul, the world, and God. In the "Transcendental Dialectic," where Kant investigates the natural antinomies which arise from the transcendental ideas, he concludes that, unlike the capacity for understanding whereby immanent objects are cognized under rules, the real use of pure reason in theoretical philosophy *cannot consist of constitutive cognitions* of the unconditioned.

In the Appendix to the "Transcendental Dialectic," Kant concludes that the only justified application of pure theoretical reason in its real use is *regulative*:

> The hypothetical use of reason, on the basis of ideas as problematic concepts, is not properly **constitutive**, that is, not such that if one judges in all strictness the truth of the universal rule assumed as a hypothesis thereby follows; for how is one to know all possible consequences, which would prove the universality of the assumed principle if they followed from it? *Rather, this use of reason is only regulative, bringing unity into particular cognitions as far as possible* and thereby **approximating** the rule to universality (A647/B675).[71]

A few lines above this paragraph, Kant defines the hypothetical use of reason, in this context, as a method whereby particular cases are subsumed under an unconditioned rule, the universality of which is uncertain, to see if the particular cases do justifiably follow. Without a contradictory case, the universal assumption of the rule can be inferred, though the rule's certainty is only problematic, not apodictic (A646-647/B674-675). Thus, theoretical reason cannot provide certainty in the correspondence between the transcendental ideas and real entities. However, the use of these ideas in theoretical reason are justified in order to subsume all conditioned cognitions under universal principles.

I will not delve into the myriad ways that Kant's notion of practical reason changed from the *First Critique* to the *Groundwork*, but suffice it here to say that in the "Canon" of the *First Critique*, Kant suggests that reason does have a practical use (A796-797/B824-825). The important development in the *Groundwork* is Kant's revolutionary presentation of practical reason's fundamental principle as a principle of autonomy (IV, 440). In the *Second Critique*, Kant will ultimately investigate how it is that

[71] Italics added for emphasis. The bold type is Kant's original emphasis.

pure reason can be practical. In its practical use, reason, as one of the higher order capacities, is structured by the principle of autonomy. In its theoretical use, the capacity for reason only has a constitutive function when applied to other capacities, such as the *understanding* (*Verstand*). Pure reason sequestered to an inquiry into the conditions of theoretical cognitions has a merely regulative use. However, in the practical use of this higher order faculty, pure reason does have a constitutive function, which is the autonomous legislation of the moral law.

Like all other capacities, the use of normative statements are justified with regard to the capacity for pure practical reason because the capacity can fail. But how can this be? The moral law is pure reason determining its own activity autonomously, and as such, there is no dialectic, or antinomy, of pure practical reason in cases of moral judgement (V, 3–14). A failure of practical reason, then, cannot come from within itself. Practical failure must come from the interference of another capacity with the proper functioning of pure reason. Humans are not fully rational beings. They are *sensibly affected rational beings*. The capacity for pure practical reason, then, faces interference from desires and incentives of inclination. In contrast to the SCC theory, the alternative interpretation of Kant suggests that incentives of inclination are not reasons, nor are they the raw materials of the world that can be constructed into reasons. Rather, these incentives of inclination are purely psychologically causal mechanisms which suggest a course of action that promises pleasure. Incentives of inclination, then, can interfere with the capacity for determination by the rational will. Thus, the capacity for pure practical reason can fail, which here means that the rational will is not fully determined to perform an action in accord with the moral law. It is precisely because humans are sensibly affected rational beings, and not purely rational, that they experience the moral law and the principle of autonomy as an imperative. It is because the human will is not fully determined, either by the moral law or the law of nature, that we can justify statements about *what one should do*.

Kant clarifies this exact point in the *Second Critique*'s "Analytic of Pure Practical Reason," Theorem IV, in which he writes, "Thus the moral law expresses nothing other than the *autonomy* of pure practical reason, that is, freedom, and this is itself the formal condition of all maxims, under which alone they can accord with the supreme practical

law" (V, 33). The fundamental principle of practical reason, the principle of autonomy, is constitutive of *pure* practical willing. Kant does not endorse the claim that the categorical imperative constitutes every act of rational choice.

Rational beings make judgments of practical reason by subjecting maxims to an internal principle which determines the correct functioning of practical reason. The relevant passage in which Kant makes it clear that the principle of autonomy is internal to the rational being's consciousness of morality is the famous passage from the *Second Critique* concerning the fact of reason (*Faktum der Vernunft*). In this passage, Kant argues that pure reason's law-like form is present within consciousness as a synthetic a priori principle (V, 31).

Kantian constitutivism will fail in reasoning a skeptic into the moral point of view for two reasons. First, as has already been shown, the fact of reason requires us to accept a stronger notion of rationality than simple instrumentalism. Second, Kant makes another controversial claim when he writes that freedom is the condition of the moral law. In a famous footnote early in the *Second Critique*, Kant claims that freedom is the *ratio essendi* of the moral law (V, 5). Kant then qualifies the *ratio essendi* claim by stipulating that the moral law is the *ratio cognoscendi* of freedom (V, 5). Notice that, while first-personal, this is a much stronger notion of moralized freedom than Korsgaard's notion of a capacity to reflect and act.

Through cognition of the moral law, rational beings become aware of the conditions which allow this cognition's possibility. The condition of the moral law is freedom. The faculty of reason seeks to unify aggregates into a rational science (*Wissenschaft*) (A832/B860).[72] For this purpose, reason requires a formal principle which can provide the conditions for cognizing each part a priori in a system connected by laws (A645/B673). Freedom, as the *ratio essendi*, serves to *ground the system of moral cognition*. We are rationally justified in presupposing freedom and categorical authority because moral consciousness presents it to us as *Faktum*, not because we identify as agents in our reflectively endorsed actions.

Assume, then, that we accept what Kant says here about the constitutive form of the pure practical will, the fact of reason, and freedom. How is it that rational beings who

[72] Compare Schafer (2020), p. 664.

are not purely rational actually do will maxims in accord with the moral law? Kant's answer hinges upon the distinction he makes in the *Metaphysics of Morals* between the two faculties of desire, *Wille* and *Willkür*. For Kant, *Willkür* is the capacity for choice, or the lower faculty (VI, 213). The higher faculty of desire is the ability of self-determination in the will, or *Wille* (VI, 213).

Reason has the capacity to determine itself from its own self-conscious activity (*Wille*).[73] Reason also has the capacity to choose to bring about an object of desire (*Willkür*), either in accordance with the law of pure practical reason or in accordance with self-love. We can fail to aim at the categorical imperative in action, but we cannot fail to aim at the categorical imperative in judgements guided by the principle of pure practical reason. The justification of moral autonomy is not something we appeal to independently of moral cognition (*ratio cognoscendi*) and its characteristic activity. Kant's moral theory is not attempting to argue with a moral skeptic, then, because there is no need to provide an independent justification for engaging in an activity that rationality engages in of its own accord.

While it is true that incentives of inclination have just as much claim to the human will as the incentive of respect for the moral law (*Achtung*), Kant's point is that we do identify with the intelligible part of the self when we recognize our consciousness of the moral law (IV, 452-453). It is only through our intelligible nature and the practical standpoint that we find a constitutive cognition of pure reason. If we define *the moral problem* as a worry that our true moral judgements fail to motivate us at all, then some form of the *practical cognitivism view* seems well-suited to answer this puzzle. Although the SCC theory proposes its *agential cognitivism view* in order to answer the moral problem (Korsgaard, 2008, p. 302-326), it has been shown that the SCC theory has trouble grounding anything like the categorical moral norms needed for a Kantian theory.

To get around the moral problem, some philosophers (Enoch 2006, 2009; FitzPatrick 2005, 2013). have proposed that the agential cognitivism assumption should be replaced with a robust, non-natural realist account of moral reasons. Instead of treating ethical reasons as internal to agency, these philosophers have suggested that

[73] For reason's self-conscious activity, see Engstrom (2009), p.98-104, and Schafer (2020).

the agent should take up reasons into his desire set when engaging in practical reasoning. This proposal is meant to get around the motivational horn of the moral problem dilemma, while the truth-aptitude horn of the dilemma is handled by virtue of the realist account of reasons. However, if we are searching for a universal and necessary account of moral truth, the robust realist account is not the only viable answer. In response to the robust realist, a Kantian can draw upon the alternative interpretation that I have sketched above in order to present a modified version of the practical cognitivism view. The reformulated assumption can be set forth in the following way:

> **Autonomous Capacity Cognitivism:** There are several rational capacities, subject to distinct norms, which each contribute to the rational being's overall functioning. This includes such capacities as the understanding, judgement, and reason. All capacities individually meeting their own internal norms will provide a standard of virtue for the rational functioning of an agent. [74]

Insofar as the capacities are distinct, the reasons generated from these capacities are also distinct in kind, such that a reason to be prudential, for example, cannot serve as a moral reason in the relevant respect if these reasons come from different sources. If we want to frame this view in terms of agency, then the proper functioning of agency relies on the norms which apply to several distinct capacities. Moral agency, and not merely an instrumental notion like efficient agency, would never be escapable by simply neglecting one capacity in favor of another. Say that an agent decides to reject his moral personhood. This is merely a situation in which we are justified in saying that the moral agent *should* act in accord with practical reason. It is a practical assumption that the capacity for moral reason has preeminent authority because this assumption alone is sufficient to justify the type of moral experience that Kant describes in the fact of reason passage.

Where Korsgaard is likely to speak of a single capacity—the capacity to be a self-conscious rational agent—it is better, by contrast, to speak of several capacities, each of which have their own rationally structured function. These capacities can work together for the purpose of *full rationality*, or these capacities can interfere with each

[74] This view is similar to the reading of Kant proposed in (Schafer, 2024).

other, such that *full rationality fails* and normative statements are justified with regard to the function of each. Insofar as a conscious being possesses a rational capacity, normative discourse justifiably applies to this being. Insofar as a conscious being possesses the capacity for practical reason, this being possesses moral personhood.

Normativity comes in degrees. Kant hints at as much when, in his *Religion within the Boundaries of Mere Reason*, he typologizes the three different predispositions which coexist within human nature (VI, 26). These predispositions are *animality*, or the predisposition to mechanical self-love; *humanity*, or the predisposition to reason instrumentally about how to gain the ends suggested by culture; and *personality*, or the predisposition to respect for the moral law (VI, 26-28). Kant presents these as predispositions to good. In themselves, there is nothing evil about the exercise of these predispositions. And, the proper function of these predispositions, governed by principles of reason, lead to the good (VI, 28; Wood, 42).

Kant emphasizes that the predisposition to animality does not require reason. Animals lack a rational will (*Wille*), and as such, they merely represent and aim at ends, such as the ends of self-preservation and species-preservation, which Kant describes in the *Metaphysics of Morals* (VI, 420). Mere humanity—that is, the technical sense at use in the predisposition passage—consists in the activity of rationality for the purpose of achieving certain ends.[75] This conception of humanity is partial and refers only to the concept of an empirically rational being, viewed from the standpoint of the sensible world.[76] In the *Religion*, Kant states that, "[The predisposition to humanity] is rooted in a reason which is indeed practical, but only as subservient to other incentives" (VI, 28).[77] However, humanity, broadly construed, is not only empirically rational, but practically rational, in the strictly moral sense of the term, by virtue of the capacity to condition the activity of reason from the principle of autonomy. Thus, moral personality—or, as Kant refers to it in the *Second Critique*, "freedom and independence from the mechanism of

[75] Compare Kant's comments on the *assertoric* nature of happiness as an end in *Groundwork* II (IV, 415-416): "One must present it as necessary not merely to some uncertain, merely possible purpose, but to a purpose that one can presuppose safely and a priori in every human being, *because it belongs to his essence*." Emphasis is my own.

[76] On the human being as part of the sensible world, see also (V, 86-87).

[77] For the instrumentalist reading of the predisposition to humanity, see also, Allen Wood's (2014) "The Evil in Human Nature," esp. pp. 41–42.

the whole of nature, regarded nevertheless as also a capacity of a being subject to special laws" (V, 87)—is the basis for "many expressions that indicate the worth of objects according to moral ideas" (V,87).[78]

In this way, the authority of a norm only holds insofar as a being has the capacity which grounds the norm. If, like Kant, one takes there to be a capacity for practical reason, and takes this capacity to be separate from the general capacity to set ends for oneself, then the worry about escaping moral normativity becomes less pressing. When a being acts from self-love, then he is merely acting on the inclinations which are imparted to him by virtue of his animality. As Kant states explicitly, animality by itself does not require practical rationality.[79] But, once rationality is allowed to structure the activity of practical reasoning through the principle of autonomy, then practical judgements must be inescapable, in the sense that the principle of autonomy inescapably determines the moral law. Of course, to say this much is not to claim that freedom of choice inescapably aims at the principle of autonomy. As has already been shown, the moral law is constitutive of *Wille*, not free choice (*Willkür*).

4. Conclusion

In this chapter, I have raised the insufficient authority objection and argued that practical identities that arise from the ground of rational agency are ultimately only regulative standards. As such, the normativity that arises from these identities is escapable. The SCC theory does not provide a ground sufficient to account for Kantian categorical moral authority. In contrast, I have set forth my own interpretation of Kant's constitutivism to see how Kant attempts to ground categorical moral authority. I found that Kant separates the capacity to make moral judgements from other non-moral rational capacities. I also found that Kant does not attempt to account for both projects of normative explanation and moral justification by appealing to one, unified capacity. Instead, Kant takes that there is a capacity to make judgements of pure practical reason, and the fact that this capacity interacts with other capacities produces moral

[78] As an example of one such moralized expression, Kant presents the FH: one should always treat humans as ends, and never as mere means (V, 87).

[79] For similar reasons, Tamar Schapiro refers to self-love as a "sham principle" that "purports to play the role of a principle, even though it is unfit to do the job that principles are supposed to do" (2021, pp.144).

normativity. The justification of this capacity's authority is a separate project that is not independent of pure reason's autonomous activity.

Chapter Five

1. Some Prescriptions

In this concluding chapter, I will make a few concluding remarks about Kantian constitutivism and how the SCC theory might accommodate my findings in order to get around the insufficient authority objection. Before I proceed, I will provide my prescriptions as a list of points for those who subscribe to the SCC theory to consider. In the following sections, I will describe why I take these prescriptions to be important, and how I take them to get around the problems I have raised.

1. The reason that an agent has to achieve some non-morally desired end should not be for the function or purpose of unifying the agent as fully rational.
2. The SCC theory should emphasize what I have termed *autonomous capacity cognitivism*. This assumption goes some way to answering how reasons can have different functions.
3. Constitutivists should not equate the projects of normative explanation and the justification of moral authority.
4. Finally, and related to point three above, constitutivism alone cannot provide the notion of categorical moral authority at issue in Kantian philosophy.

In what follows, I will conclude by outlining the above four points in more detail. I take it that this material is all a summary of the conclusions reached from the foregoing arguments.

2. Autonomous Capacity Cognitivism

This section will largely look to the idea of framing Kantian constitutivism in the terms of *autonomous capacity constitutivism* as I outlined it in the preceding chapter. In particular, I apply this new assumption to the normative problem and the Henry VIII case to see what conclusion it generates.

There are a few distinct points that *autonomous capacity cognitivism* emphasizes. First, rational beings have several distinct rational capacities, such that these capacities can work in tandem for *full rationality*, or these capacities can interfere with each other such that normative reason arises for their proper function. Second, the virtuous

functioning of a capacity comes in degrees. For example, one capacity interfering with the rational function of another capacity will render this second capacity less than fully rational and also less than fully virtuous.

If Kantian constitutivists want to emphasize agency,[80] the proper functioning of the agent relies on the norms which apply to several distinct capacities, such that skepticism concerning the authority of one capacity cannot impugn the viability of the theory's account of normativity. Under this new interpretation, all reasons do come from the same source, insofar as this source is rationality. However, rationality is subject to different uses and subjects of inquiry. As such, the different functions of rationality, which are determined by each of the separate capacities, allow for reasons of different kinds and uses. Epistemic reasons might come from a capacity for theoretical judgements, while practical reasons, in the Kantian use of that term, will come from a capacity to make moral judgements.

In order to see how the new interpretation, along with the assumption of *autonomous capacity cognitivism*, fares with regard to the problem of the normative, we can return to the Henry VIII case. In this case, when the courtier is faced with threats from the king, he *must* choose to determine his course of action because he is a rational being with the power for free choice. But the freedom of choice is not constitutive of the moral law. The principle of autonomy dictates that the courtier *should* do the moral action and deny the king's threats. From the practical point of view, the courtier *should* endorse a maxim that incorporates an end suggested by the principle of autonomy because this is *more indicative of the rational being's true self* than an end that comes from inclination. To deny autonomy to oneself is to objectify oneself as causally determined and heteronomous.

Any other form of normativity in this case would be non-moral normativity, and it would not come from the courtier's capacity for pure practical reason. If, like the SCC theory, we want to frame incentives of inclination as possible data from the world which could be constructed into a reason, then these reasons would merely suggest a course of action sufficient to bring about some desired end. For example, this kind of

[80] I will remain agnostic on what a Kantian theory of agency should look like.

normativity would suggest how it is that the rational being could achieve non-moral ends, such as pleasure or prudence.

One set of reasons suggests doing the moral action, and another set of reasons suggests instrumental efficiency. These reasons are generated by separate capacities, and as such, a person might be said to function well with respect to one capacity and yet fall short of the virtuous exercise of the other capacity. Choosing to will a private reason—where this refers to a reason not in accord with moral judgement—is not escaping the constitutive function of agency so much as simply exercising one capacity while neglecting the other. On this view, if an agent attempts to neglect the constraints set by pure practical reason, then the agent is simply behaving poorly with respect to the function of moral virtue. This would constitute a case of self-deception or willfully bad action.

In more traditional Kantian terms, prudential reasoning and hypothetical imperatives are merely a species of theoretical reasoning. On this interpretation, hypothetical imperative provide knowledge of how to execute a series of actions such that the desired end goal could be brought about.[81] Furthermore, the orthodox Kantian would say that incentives of inclination are ultimately non-cognitive, and thus, normativity cannot come from these incentives, whether constructed into a reason or otherwise. Instead, non-moral normativity, in this view, has more to do with the ability to comprehend than with the ability to act rationally.[82] In orthodox Kantian terms, then, non-moral normativity stems from the capacity for understanding and the use of theoretical reason. Similar to the case of a reasoner following a rule of inference, this view would simply equate what the courtier has non-moral reason to do with an end of epistemic virtue.

In effect, this section has answered points one and two above. On the first point—dispelling the new bootstrapping worry—the agent's emotions are not entirely governed by rational deliberation. As Kantians, we should either allow that non-moral normativity can be constructed from incentives of inclination, in which case the reason to feel a certain way cannot be entirely governed by the function of moral virtue; or, we

[81] See (Timmermann, 2022).
[82] See (Schafer, 2019b, 2023).

should stand with the traditional Kant interpretation and suggest that non-moral normativity only aims at epistemic virtue, in which case the agent's emotions are the material for self-comprehension.

On the second point above—the theoretical adequacy of the *autonomous capacities cognitivism* view—we have seen that this does generate the correct conclusion in the Henry VIII case. Another benefit of this view is that it presents a clear example of normative explanation. This leads to the third and fourth prescriptions, which I detail in the next section.

3. Explanation, Justification, and the Limits of Identity

With the above conclusion, I will now provide an answer for points three and four above. In effect, this section will show the limits of constitutivism as a metaethical theory. The assumption that one should act from the principle of autonomy, that this principle has categorical moral authority, is not above suspicion. In general, the theory, as I have presented it, can do nothing to argue the skeptic into bowing before morality's categorical commands. But, Kant was never attempting to respond to the moral skeptic in the first place. His theory is structured as a justification of the pre-theoretical conception of morality. In the *Groundwork*, Kant takes the truth of morality for granted, even from the beginning.

Constitutivists cannot attempt to answer moral skepticism by conflating a *normative explanation* of rationality's function with a *justification* of the moral capacity's predominant authority. The skeptic asks, why should I identity with moral agency? In response, it is not enough to justify moral authority by suggesting that the skeptic must reflect and act. If the normative explanation and justification are conflated—as the SCC theory attempts to do—then both projects succeed or fall together.

Through the interpretation of Kant that I have proposed, I have already shown how constitutivism can keep these two projects distinct, but I will belabor this point here in order to make it explicit. How can we answer the question of normative explanation? Kant answers this question by appealing to distinct capacities which are functionally structured by the synthetic a priori principles that reason represents to itself. A rational capacity, left to its own accord, will perform its representative function because the

principle that governs its behavior is properly constitutive of its function. However, some capacities can conflict with each other such that these capacities malfunction.

Under this interpretation, we can define the categorical imperative—the moral law presented to the rational will as a categorical norm—as the capacity for pure practical reason (*Wille*) conflicting with a sensibly conditioned capacity for free choice (*Willkür*). We can provide a description of rationality's function by appealing to the guiding principles which reason represents to itself. Normativity arises with respect to these capacities when the capacity fails in the virtuous execution of its function.

The above was an answer to the question of normative explanation. Normativity is explained by appealing to rationally structured capacities with the possibility of failure. But, there is still the question of the moral capacity's authority. What justifies the categorical authority of the capacity for pure practical reason? Why is the moral law's authority binding such that it can defeat immoral maxims? Notably, this is distinct from the question, *why be moral at all?* Kant first attempts to answer this question in *Groundwork* III where he essays a transcendental deduction of the categorical imperative. Here, one can see the temptation to frame the justification of moral authority in terms of identity. Kant describes the intelligible nature of humanity as the "true self" as opposed to the casually determined self of the sensible world (IV, 461). Ultimately, Kant's answer is that pure practical reason is the capacity for the rational being to be *autonomous*.

In *Groundwork* III, identity is playing a secondary role in the justification of pure practical reason's categorical authority. Rational beings identify with autonomy as the true self because this is the only constitutive use of *pure reason*. This assumption is part of Kant's transcendental idealism. In the realm of metaethics, we can frame this as Kant's *metaethical idealism*, an assumption that autonomy is the synthetic a priori principle which serves as the only constitutive use of pure reason. As a self-given principle of reason, autonomy is the only possible cognition of the self that is not

determined by the causal laws of the phenomenal realm. Therefore, in *Groundwork* III, the justification of morality's categorical authority relies on Kant's idealist assumptions.[83]

Metaethical idealism is, of course, in stark contrast to the pragmatism of the SCC theory. One benefit of SCC theory is its metaphysically parsimonious account of normativity. Kant eventually gives a response to the question of moral authority's justification which plays down the aspects of transcendental idealism emphasized in *Groundwork* III. In the *Second Critique*, Kant provides the fact of reason as the justification of morality's categorical authority. Unlike Korsgaard's justification of freedom in *Sources*, the fact of reason is not the freedom to reflect and act, but the freedom of rationality generating the moral law of its own accord. As I discussed in Chapter 4, the fact of reason is a justification that starts from a moral experience that already entails autonomy as an implicit feature. Kant's fact of reason takes moral experience at face value. It is not, as Korsgaard would have it, the effect of turning one's reflective attention inward, but the spontaneous cognition of one's own autonomy.

If as I have been arguing, constitutivists never should have equated the projects of normative explanation and justification, then Kantian constitutivism should look more like the interpretation of Kant that I have put forward. The normative explanation still relies on the rational structure of certain capacities that can fail in their characteristic exercise. But, the justification of morality's categorical authority must provide a separate answer. One such response is Kant's answer in *Groundwork* III, which emphasizes transcendental idealism. Another such response is Kant's answer in the *Second Critique*, which emphasizes pure reason's spontaneous exercise of moral autonomy.

While the prospect of separating normative explanation from justification does answer the insufficient authority objection, this response makes clear the limits of constitutivism as a metaethical theory. If we provide a constitutive explanation of moral normativity, then the justification of morality's categorical authority must appeal to something other than the constitutive explanation. Constitutivism, by itself, *can* provide

[83] For an interpretation that suggests Kant attempts a theoretical proof, see Dieter Henrich's classic "Identity and Objectivity: An Inquiry into Kant's Transcendental Deduction" (1994, pp. 123–210). For an attempt to read the deduction as a practical proof, see (Rauscher, 2009) "Freedom and reason in *Groundwork* III."

an account of a normatively structured capacity for moral reasoning. However, constitutivism, by itself, *cannot* provide an account of categorical moral authority.

Kant saw the difficulty of providing an adequate response to the justification question in the *Groundwork*. In the last sentence of that work, Kant writes, "And thus we do not indeed comprehend the practical unconditional necessity of the moral imperative, yet we do comprehend its *incomprehensibility*, and this is all that can reasonably be required of a philosophy that in its principles strives up to the boundary of human reason" (IV, 463). While these limits might attenuate the remit of inquiry, there is nothing in these theoretical limits which render a modified Kantian constitutivism unworkable. Metaethical realists, for example, will face similar hurdles in attempting to justify the practical nature of a non-natural reason that exists in the world, apart from the agent.[84] Once we resolve the dilemma, we see that constitutivism can be a helpful tool in the fraught debates surrounding the sources of normativity, but by itself, constitutivism will not be the Archimedean point by which philosophers move the world.

4. Concluding Remarks

I take it that the conclusions found here will be of interest to Kantians but as well to any philosopher interested in practical reasoning, metaethics, and moral psychology. In order to even have this debate, it seems that we needed first to have in hand the tools which Kant's theory provides. Once again, Kant's Tower of Babel analogy proves true: "It turned out, of course, that although we had in mind a tower that would reach the heavens, the supply of material sufficed only for a dwelling that was just roomy enough for our business on the plane of experience and high enough to survey it" (A707/B735). In that passage, Kant is speaking about the limits of reason in its speculative use, but it could just as well warn about the limits of agency.

[84] Schafer (2023, pp. 113): "But note that [the shmagency point] is a limitation that applies to any conceivable meta-ethical view whatsoever."

Bibliography

Allison, Henry E. (2021). “Reason (*Vernunft*),” in *The Cambridge Kant Lexicon*, edited by Julian Wuerth, 361–371. Cambridge University Press.

Bader, Ralf M. (2023). “The Dignity of Humanity,” in *Rethinking the Value of Humanity*, edited by Sarah Buss and L. Nandi Theunissen, 153–180. Oxford University Press.

Broome, John. (1999). “Normative Requirements.” *Ratio*, 12, 398–419.

————————. (2020). “Rationality versus Normativity.” *Australasian Philosophical Review*, vol. 4 (4), 293–311.

Brunero, John. (2004). “Korsgaard on Motivational Skepticism.” *The Journal of Value Inquiry*, vol. 38, 253–364.

————————. (2005). “Instrumental Rationality and Carroll’s Tortoise.” *Ethical Theory and Moral Practice*, vol. 8, 557–569.

————————(2021). “Ambivalence, Incoherence, and Self-Governance,” in *The Philosophy and Psychology of Ambivalence: Being of Two Minds*, edited by D. Gatzia and B. Brogaard. Routledge.

Cohen, Alix. (2017). “Kant on Emotions, Feelings, and Affectivity.” In *The Palgrave Kant Handbook*, ed. by M.C. Altman. 665–681. Palgrave Macmillan.

————————. (2018). “Kant on Moral Feelings, Moral Desires and the Cultivation of Virtue.” In *Begehren/Desire*, ed. by Sally Sedgwick and Dina Edmundts. 3–18. De Gruyter.

Copp, David. (2000). “Korsgaard on Normativity, Identity, and the Ground of Obligation,” in *Rationalität, Realismus, Revision/ Rationality, Realism, Revision*, edited by Julian Nida-Rümelin, 572–581. De Gruyter.

Dasgupta, Shamik. (2017). “Constitutive Explanation.” *Philosophical Issues: A Supplement to NOÛS*, vol. 27, 74–97.

Engstrom, Stephen. (2009). *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Harvard University Press.

Enoch, David. (2006). “Agency, Shmagency: Why Normativity Won’t Come from What Is Constitutive of Action.” *Philosophical Review*, 115, 2, 169–198.

————————. (2009. “Can there be a global, interesting, coherent constructivism about practical reason?” *Philosophical Explorations*, vol. 12 (3), 319–339.

Falk, W.D. (1947/8). “ ‘Ought’ and Motivation.” *Proceedings of the Aristotelian Society*, vol. 48, 111–138.

FitzPatrick, William J. (2005). “The Practical Turn in Ethical Theory: Korsgaard’s Constructivism, Realism, and the Nature of Normativity.” *Ethics*, 115 (4), 651–691.

————. (2013). “How not to be an ethical constructivist: A critique of Korsgaard’s neo-Kantian constitutivism.” In *Constructivism in Ethics*, ed. by Carla Bagnoli. 41–62. Cambridge University Press.

Fix, Jeremy David. (2020). “The Error Condition.” *Canadian Journal of Philosophy*, vol. 50, (1), 34–48.

————. (2020). “The Instrumental Rule.” *Journal of the American Philosophical Association*, vol. 6, (4), 444–462.

————. (2021). “Two Sorts of Constitutivism.” *Analytic Philosophy*, vol. 62, (1), 1–20.

————. (2023). “Grounds of Goodness.” *Journal of Philosophy*, vol. 120, (7), 368–391.

Frankfurt, Harry G. (1988). *The Importance of What We Care about: Philosophical Essays*, Cambridge University Press.

Greenspan, P.S. (1975). “Conditional Oughts and Hypothetical Imperatives.” *The Journal of Philosophy*, vol. 72, (10), 259–276.

Guillot, Marie and Lucy O’Brien. (2022). “Self Matters.” *Ergo: An Open Access Journal of Philosophy*, vol. 9 (28), 728–754.

Gunnarsson, Logi. (2014). “In Defense of Ambivalence and Alienation.” *Ethical Theory and Moral Practice*, vol. 17, 13–26.

Hare, RM. (1952/1964). *The Language of Morals*. Oxford University Press.

Henrich, Dieter. (1994). “Identity and Objectivity: An Inquiry into Kant’s Transcendental Deduction,” translated by Jeffrey Edwards, in *The Unity of Reason: Essays on Kant’s Philosophy*, edited by Richard L. Velkley, 123–210. Harvard University Press.

Hill, Thomas E. (1992). “The Hypothetical Imperative,” in *Dignity and Practical Reason in Kant’s Moral Theory*, 17–38. Cornell University Press.

Hills, Alison. (2008). “Kantian Value Realism.” *Ratio*, vol. 21, (2), 182–200.

————. (2015). “Cogntivism about Moral Judgement,” in *Oxford Studies in Metaethics*, vol. 10, edited by Russ Shafer-Landau, 1–25. Oxford University Press.

Kant, Immanuel. *Practical Philosophy*, translated and edited by Mary Gregor, for *The Cambridge Edition of the Works of Immanuel Kant* (1996). Cambridge University.

————.*Critique of Pure Reason*, translated and edited by Paul Guyer and Allen W. Wood, for *The Cambridge Edition of the Works of Immanuel Kant* (1998). Cambridge University Press.

————.*Critique of the Power of Judgement*, edited by Paul Guyer, translated by Paul Guyer and Eric Matthews, for *The Cambridge Edition of the Works of Immanuel Kant* (2000). Cambridge University Press.

————. *Notes and Fragments*, translated and edited by Paul Guyer, translated by Curtis Bowman and Frederick Rauscher, for *The Cambridge Edition of the Works of Immanuel Kant* (2005). Cambridge University Press.

————. *Religion within the Boundaries of Mere Reason: And Other Writings*, translated and edited by Allen Wood and George Di Giovanni (1998). Cambridge University Press.

————.*Groundwork of the Metaphysics of Morals*, edited and translated by Mary Gregor and Jens Timmermann (2011). Cambridge University Press.

————.(1724-1804). Immanuel Kant: *Gesammelte Schriften (Akademie-Ausgabe)*, vols. I-XXIII, edited by Königlich Preussischen Akademie der Wissenschaften Berlin: G. Reimer (1902). De Gruyter (1922).

Katsafanas, Paul. (2018). “Constitutivism about Practical Reasons.” In *The Oxford Handbook of Reasons and Normativity,* ed. by Daniel Star. 367–391. Oxford University Press.

Kolodny, Niko. (2005). “Why be Rational?” *Mind*, vol. 114, (455), 509–563.

Korsgaard, Christine M. (1995). “Rawls and Kant: On the Primacy of the Practical.” *Proceedings of the Eighth International Kant Congress*, vol. 1, 1165–1173.

————(1996a). *Creating the Kingdom of Ends*. Cambridge University Press.

————. (1996b). *The Sources of Normativity*. Cambridge University Press.

————.(1998). “Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind.” *Ethics*, vol. 109, 49–66.

————. (2008). *The Constitution of Agency: Essays of Practical Reason and Moral Psychology*. Oxford University Press.

————. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.

————————.(2013). “The Relational Nature of the Good,” in *Oxford Studies in Metaethics*, vol. 8, edited by Russ Shafer-Landau, 1–26. Oxford University Press.

————————.(2014). “On Having a Good.” *The Royal Institute of Philosophy*, vol. 89, 405–429.

————————.(2019). “Constitutivism and the Virtues.” *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, vol. 22 (2), 98–116.

————————.(2021). “Valuing our Humanity,” in *Respect: Philosophical Essays*, ed. by Richard Dean and Oliver Sensen, 171–191. Oxford University Press.

Langton, Rae. (2007). “Objective and Unconditioned Value.” *Philosophical Review*, Vol. 116, (2), 157–185.

Luce, Duncan R. and Howard Raiffa. *Games and Decisions: Introduction and Critical Survey*. Dover Publications.

Moran, Richard. (2002). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.

————————. (2022). “Self-Consciousness and Self-Division in Moral Psychology,” in *Normativity and Agency: Themes from the Philosophy of Christine M. Korsgaard*, edited by Tamar Schapiro, Kyla Ebels-Duggan, and Sharon Street, 95–125. Oxford University Press.

O’Neill, Onora. (1989). *Constructions of Reason: Explorations of Kant’s Practical Philosophy*. Cambridge University Press.

————————. (2006). “Constructivism in Rawls and Kant,” in *The Cambridge Companion to Rawls*, edited by Samuel Freeman, 347–367. Cambridge University Press.

Oxford English Dictionary, s.v. “ambivalence (n.), sense 1.b,” September 2024, https://doi.org/10.1093/OED/3055940960.

Rauscher, Frederick. (2002). “Kant’s Moral Anti-Realism.” *Journal of the History of Philosophy*, vol. 40, (4), 477–499.

————————. (2009). “Freedom and reason in Groundwork III,” in *Kant’s Groundwork of the Metaphysics of Morals: A Critical Guide*, edited by Jens Timmermann, 203–223. Cambridge University Press.

Rawls, John. (1980). “Kantian Constructivism in Moral Theory.” *The Journal of Philosophy*, vol. 77 (9), 515–572.

————————.(1971/1999). *A Theory of Justice*. Belknap, Harvard University Press.

Reath, Andrews. (2010). “Formal principles and the form of a law.” In *Kant’s Critique of Practical Reason: A Critical Guide*, ed. by Andrews Reath and Jens Timmermann. 31–54. Cambridge University Press.

————. (2013). “Formal Approaches to Kant’s Formula of Humanity.” In *Kant on Practical Justification: Interpretive Essays*, ed. by Mark Timmons and Sorin Baiasu. 201–228. Oxford University Press.

————. (2022). “Kantian Constructivism and Kantian Constitutivism: Some Reflections.” *Kant Yearbook*. 14 (1), 45–69.

Samuel, Jack. (2023). “Toward a Post-Kantian Constructivism.” *Ergo: An Open Access Journal of Philosophy*, vol. 9, (53), 1449–1484.

Schafer, Karl. (2015a). “Realism and Constructivism in Kantian Metaethics (1): Realism and Constructivism in a Kantian Context.” *Philosophy Compass*. 690–701.

————. (2015b). “Realism and Constructivism in Kantian Metaethics (2): The Kantian Conception of Rationality and Rationalist Constructivism.” *Philosophy Compass*. 702–713.

————. (2018). “Constitutivism about Reasons: Autonomy and Understanding.” In *The Many Moral Rationalisms*, ed. by Karen Jones and Francois Schroeter. 70–90. Oxford University Press.

————. (2019a). “Kant: Constitutivism as Capacities-First Philosophy.” *Philosophical Explorations*. 22 (2), 177–193.

————. (2019b). “Rationality as the Capacity for Understanding.” *NOÛS*. 53 (3), 639-663.

————. (2020). “Transcendental Philosophy as Capacities-First Philosophy.” *Philosophy and Phenomenological Research* vol. 103, 661–686.

————. (2023). *Kant’s Reason: The Unity of Reason and the Limits of Comprehension in Kant*. Oxford University Press.

Schapiro, Tamar. (2021). *Feeling Like It: A Theory of Inclination and Will*. Oxford University Press.

Sellars, Wilfrid. (1956/1997). *Empiricism and the Philosophy of Mind*. Harvard University Press.

Sensen, Oliver. (2009). “Kant’s Conception of Inner Value.” *European Journal of Philosophy*, vol. 19 (2), 262–280.

————————.(2017). “Kant’s Constitutivism.” In *Realism and Antirealism in Kant’s Moral Philosophy: New Essays*, ed. by Robinson dos Santos and Elke Elisabeth Schmidt. 197–222. De Gruyter.

————————.(2022). “Kant’s Value Prescriptivism,” in *Kant’s Theory of Value*, edited by Christoph Horn and Robinson dos Santos, 23–40. De Gruyter.

Shakespeare, William. *The Tragedy of Hamlet: Prince of Denmark*, edited by Barbara A. Mowat and Paul Werstine. The Folger Shakespeare. https://www.folger.edu/explore/shakespeares-works/hamlet/read/.

Smith, Michael. (1994). *The Moral Problem*. Blackwell.

————————. (2007). “Is There a Nexus Between Reasons and Rationality?” in *Moral Psychology*, *Poznań Studies in the Philosophy of the Sciences and the Humanities*, vol. 94, edited by Sergio Tenenbaum, 278–298.

————————. (2013). “A Constitutivist Theory of Reasons: Its Promise and Parts.” *LEAP*, vol. 1, 9–30.

————————. (2018). “Three Kinds of Moral Rationalism.” In *The Many Moral Rationalisms*, ed. by Karen Jones and Francois Schroeter. 48–69. Oxford University Press.

Smith, Michael, David Lewis, and Mark Johnston. (1989). “Dispositional Theories of Value.” *Aristotelian Society Supplementary Volume*, vol. 63, (1), 89–174.

Street, Sharon. (2010). “What is Constructivism in Ethics and Metaethics?.” *Philosophy Compass*. 363–384.

Sussman, David. (2003). “The Authority of Humanity.” *Ethics*, vol. 113, (2), 350–366.

Tenenbaum, Sergio. (2019). “Formalism and Constitutivism in Kantian Practical Philosophy.” *Philosophical Explorations*. 22 (2), 163–176.

Timmermann, Jens. (2005). “Good but not Required?—Assessing the Demands of Kantian Ethics.” *Journal of Moral Philosophy*, vol. 2 (1), 9–27.

————————.(2006). “Value without Regress: Kant’s ‘Formula of Humanity’ Revisited.” *European Journal of Philosophy*. 14 (1), 69–93.

————————. (2007). *Kant’s Groundwork of the Metaphysics of Morals: A Commentary*. Cambridge University Press.

———————. (2009). “Acting from duty: inclination, reason, and moral worth,” in *Kant’s Groundwork of the Metaphysics of Morals: A Critical Guide*, edited by Jens Timmermann, 45–62. Cambridge University Press.

———————. (2013). “Kantian Dilemmas? Moral Conflict in Kant’s Ethical Theory.” *Archiv für Geschichte der Philosophie*, vol. 95, (1), 36–64.

———————. (2022). *Kant’s Will at the Crossroads: An Essay on the Failings of Practical Rationality*. Oxford University Press.

Walker, Ralph C.S. (2022). *Objective Imperatives*. Oxford University Press.

Wood, Allen. (2014). “The Evil in Human Nature,” in *Kant’s Religion within the Boundaries of Mere Reason*, edited by Gordon E. Michalson, 31–57. Cambridge University Press.