GROUNDING CONSTITUTIVISM

by

Kathryn M. Lindeman

A.B., Mount Holyoke College, 2005

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial

fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Kathryn M. Lindeman

It was defended on

January 28, 2014

and approved by

Kieran Setiya, Professor of Philosophy

Robert Brandom, Distinguished Professor of Philosophy

Karl Schafer, Assistant Professor of Philosophy

Luca Ferrero, Associate Professor of Philosophy, University of Wisconsin, Milwaukee

Dissertation Director: Kieran Setiya, Professor of Philosophy

GROUNDING CONSTITUTIVISM

Kathryn M. Lindeman, PhD

University of Pittsburgh, 2014

The world is full of things that are better and worse, and we evaluate these things according to norms. However, the nature of norms and how they are related to the those better and worse things is not at all clear. Constitutive accounts of norms attempt to locate norms from as a constituent of the nature of the very individuals they evaluate. I first argue against current constitutivist accounts because of their endorsement of what I call the Threshold Commitment, which holds that a necessary condition of being a member of an evaluative kind ? a kind whose members can be evaluated qua their kind ? is satisfying some minimum number or degree of the norms of that kind. I then develop a better account of the constitutive relationship between evaluative kinds and the norms that evaluate them. By understanding evaluative kinds as constituted by functions, constitutivists can easily explain norms as those standards whose satisfaction makes individual kind-members better at performing their function. I argue that understanding constitutive functions according to an etiological account of proper function avoids endorsing the Threshold Commitment and so makes one immune to the problems with current constitutivist accounts. Thus begins a kind of proof by cases for the position that the norms governing evaluative kinds can be grounded in the functional nature of those kinds. The last chapters answer worries raised by the use of etiological function in the constitutivist account. I consider a priori and empirical worries. The a priori objections are based on our intuitions about when and how we can evaluate individuals according to norms and fears that the etiological account does violence to these intuitions. The empirical objections focus on worries that etiological functions can be found to support norms contrary to our intuitive theoretical norms.

TABLE OF CONTENTS

PREFACE

This dissertation took many years, revisions, conversations, and innumerable acts of grace to complete. In its final version, it is still maturing, but insofar as it includes any truth or insight into the dark heart of meta-ethics, it is because I was supported.

I owe a special debt of gratitude to the committee members who wore many hats in guiding me through this dissertation. It has been a joint labor, from the early stages of exposing my confusions about normative kinds with Bob Brandom, to the cup of coffee with Luca Ferrero over which the germ of this dissertation was born and the fruitful discussions that followed, to the tireless labor of Kieran Setiya to help me bring a view of my own into shape and hone it into something worth defending, to the critical support and friendship of Karl Schafer to help me shape it into a project with the scope and ambition it has today. I know people write dissertations without thoughtful, engaged, critical, and supportive advisors, but I'm still not sure how.

Those who have supported me, and my philosophical journey, during the past however-many years can never be appropriately thanked in this space, but my life and this work (and usually both) would be poorer without the following people: Dan Addison, Mikio Akagi, Bob Batterman, Justin Blank, Autumn Boyer, Eric Brown, Michael Brownstein, John Brunero, Brett Caloia, Connie Cassor, Lindsay Crawford, Billy Dunaway, Casey Doyle, Rachael Driver, Peter Dudley, Meghan Dupree, Stephen Engstrom, Sam Floyd, Chris Frey, Jenn Frey, Kim Frost, Jay Garfield, Josh Hancox, James Harold, Boris Hennig, Michael Hoffman, Kristen Inglis, Anja Jauernig, Jenn Jhun, Kelson Law, Hallie Liberto, Errol Lord, Stephen Makin, Kate Manne, John McDowell, Adam Marushak, Brooke McLane-Higgenson, Eliot Michaelson, Sam Mitchell, Shyam Nair, Jordan Offutt, Tristram Oliver-Skuse, Cameron O'Mara, Japa Pallikkathayil, Sarah Paul, James Pearson, Tom Ricketts, Ben Schultz, James

Shaw, Jamsheed Siyar, Shawn Standefer, Daniel Star, Will Starr, Catherine Stinson, Robert Steel, Preston Stovall, Greg Strom, Jonathan Surovell, Kathryn Tabb, Michael Thompson, Lynne Tirrell, Hille Paakkunainen, Katia Vavova, Milos Vuletic, Eric Wiland, Tim Willenken, Vanessa Wills, and Karen Zwier. These people have provided invaluable feedback, insight, guidance, friendship, hugs, wine, and laughs. Each in their own way, they have made my life a better one, full of philosophy, friendship, and other things worthwhile. I will be forever grateful.

In the course of writing this dissertation, I lost the person who would have been the proudest to have seen it completed. The sudden death of my father is a source of unending sadness, but the endless support of those who love me, and those who loved him, has been a grace. I want to especially thank my mother, Linda, and my brother and sister, Stuart and Laura. You generally don't get to choose your family, but I couldn't have done better. I am so grateful for their unearned love, humor, and constant reminders of what love looks like in its most rewarding and challenging forms. I did, however, get to choose Tyke Nunez. His quiet, loving support of me and my philosophical aspirations through times I have been willing to give up on both have meant the world to me.

This dissertation is for my dad, James Edward Lindeman, III, who has always been, and will always be, my biggest fan.

1.0 INTRODUCTION

The world is full of things that are better and worse, and we evaluate these things according to norms. However, the nature of norms and how they are related to the those better and worse things is not at all clear. Most salient to philosophers are practical and theoretical norms, those norms that evaluate aspects of our agentive and epistemic lives, but a unified account of norms should account for all things that can be (and can thus be judged to be) better and worse. We might reasonably ask how norms are related to their objects and what explains the connection between evaluation and norms. Weighing the viability of one explanation of this connection, called Constitutivism, is the task of this dissertation.

Constitutive accounts of norms attempt to locate norms as a constituent of the nature of the very individuals they evaluate. Most prominently, these accounts are of practical and theoretical norms, which locate these norms in the nature of agency or rationality as such. This grounding accounts for one of the main appeals of constitutivism: it explains what makes something good in terms of the nature of the kind to which it belongs, and thus promise to make norms only as mysterious as their objects. This project, of course, owes an explanation of the objects of evaluation from which constitutive norms can be derived and despite constitutivism's appeal, there is substantial controversy about whether the core constitutivist project can be successful.

The first task of my project here is to weigh prevalent objections to currently popular constitutivist accounts. I focus on the objection that constitutivists cannot account for the possibility of norm violation if norms are constitutive of their objects. This objection is taken to threaten both the violability of norms and the possibility of imperfect kind-members. The threat seems to be that if norms that are constitutive of their objects are incapable of being violated. Though this worry is pervasive throughout the literature on

this topic, I conclude that current constitutivists are able to account for the possibility of both negatively assessable (i.e. bad) individuals and violable norms. I argue that the feared consequences only arise on accounts where full satisfaction of all norms governing individual kind-members is a condition of kind-membership. I argue that no constitutivist theory is committed to this kind of constitutive account. I demonstrate that despite this easy response, continued concern about these objections is a consequence of the prevalent position of current constitutivists that the relation of constitution between norms and their objects is one of satisfaction.

Though constitutivists are able to provide violable norms, I argue current accounts are still untenable. I hold that this is because of their endorsement of what I call the Threshold Commitment, which holds that a necessary condition of being a member of an evaluative kind—a kind whose members can be evaluated qua their kind—is satisfying some minimum number or degree of the norms of that kind. This commitment can be seen in expressions like "That coffee is so bad that it isn't even coffee anymore!" In the closing sections of the second chapter, I argue that any viable constitutivist account must abandon this connection between norm satisfaction (or the disposition to satisfy norms) and being a member of an evaluative kind. Consequently, I show that the constitutivist faces the challenge of grounding norms in the nature of evaluative kinds without two of their preferred strategies: minimal norm satisfaction or the disposition to satisfy norms.

In the third chapter, I take up the challenge to develop a better account of the constitutive relationship between evaluative kinds and the norms that evaluate them. I argue that understanding evaluative kinds as constituted by functions can easily explain norms as those standards whose satisfaction makes individual kind-members better at performing their function. I argue that understanding constitutive functions according to an etiological account of proper function avoids endorsing the Threshold Commitment and so makes one immune to the problems canvased in chapter two. After all, etiological proper functions can be had by individuals that are neither able nor disposed to satisfy them. So, this account of proper function promises to provide the resources to account for the constitutive relationship between norms and evaluative kinds without running into problems that we discovered in chapter two. Thus begins a kind of proof by cases for the position that the norms governing

evaluative kinds can be grounded in the functional nature of those kinds.

Etiological proper function's first home is in the philosophy of biology, so I begin there and explain how understanding biological kinds as essentially functional kinds allows an easy explanation of biological norms. Having accounted for the biological norms as grounded in biological proper function, I extend this account, analogically, to artifactual kinds. While biological kinds have proper functions in virtue of being created with forms selected by evolutionary constraints, artifactual kinds have their functions in virtue of having been created with a form selected by a designer for that function. Norms of biological and artifactual kinds, then, can be grounded in the functional nature of the objects of evaluation. This development of metanormative constitutivism explains how function accounts for the constitutive but non-satisfaction based role of norms in kinds. It begins with the more straightforward cases of biological and artifactual kinds, and seeks to generalize from there, so that evaluation of human thought and action turn out to be an instance of something perfectly straightforward.

The next two chapters take up, in turn, two avenues of resistance to reliance on etiological function for the purpose of understanding evaluative kinds and thus pose a threat to locating the norms, particularly epistemic and theoretical norms, constitutively in functional kinds. In the fourth chapter, I consider *a priori* objections to the violence to our intuitions about when and how we can evaluate individuals according to norms in etiological function. These objections threaten to show that the constitutivist picture on offer would do violence to our intuitions about when and how we can evaluate individuals according to norms. They show that we intuitively apply norms to things that lack etiology (e.g. Swampman) and that we want to group in one evaluative kind, groups with divergent etiologies (e.g. Martians, human beings, and computers all seem potentially evaluable by the norms of rationality).

In response to these challenges, I offer an error theory of Swampnorms, and account for the possibility of judgments of swampman's parts relative to his interests, accounting for them as good *for Swampman's interests* rather than as good *qua their kind*. So, for example, Swampman's replica heart, despite not being governed by heart norms, can still nonetheless be bad *for Swampman* if it stops beating. I then respond to the worry about divergent etiologies by showing how it misunderstands the way in which evaluative kinds are

historical. Individual kind-members need not share a history to share an evaluative kind; they must share a function. An etiology is required to account for the function of each rational individual, but nothing bars the evaluative kind of rational beings from being multi-realizable. It can therefore be composed of individuals with divergent etiological accounts, including perhaps non-evolved accounts. Thus, grouping etiologically diverse individuals into one functional group is a licit move for the metanormative constitutivist.

In the fifth chapter, I consider the possibility that my view falls to empirical falsification at the hands of evolutionary psychologists. Metanormative constitutivism seems to be in a vulnerable position because of two basic worries arising from psychological results that show we are statistically bad at various reasoning tasks. These results generate two basic worries for the metanormative constitutivist. The first worry is that at least some of our intuitive theoretical norms aren't supported by an etiological function; this is a worry that the functions available to us might be too weak to do the job the metanormative theorist wants (or requires a dramatic revision of our accepted theoretical norms).

The second worry is that etiological functions can be found to support norms contrary to our intuitive theoretical norms; this is a worry that the functions available to us might be too strong and entail norms the metanormative theorist would reject. I argue that these objections misunderstand both the way functions can be had (e.g. by assuming that a form cannot have more than one proper function) or how they can be obtained (e.g. by assuming that in order for a form to gain a new function, it must first undergo some change in form). I argue against these mistaken assumptions and show how the normative violations that psychologists have demonstrated in studies are likely the result of non-optimal multi-functional mental capacities evolved over time in response to diverse adaptive problems. At the end of chapter five, having dismissed both common *a priori* and empirical objections to this role for functions in the constitutivist account of norms, I haven't proven that etiologically based constitutivism must succeed in all normative arenas, but I have solved the most pressing objections.

2.0 CONSTITUTIVISM'S PROMISE

2.1 NORMS AND EVALUATION

Philosophers worry a great deal about what makes something a reason, or a good action, or what ought to be done. Notably, those who have these worries seem to worry significantly less (or not at all) about what makes something a good thermometer, or a good knife, or the way a fern ought to grow. This is perhaps unsurprising. Thermometers, knives, ferns: these are all things that have their own experts; leave it to them to figure out the standards of artifacts and organisms. But, in another way, it is surprising. People who worry a great deal about the grounding of practical normativity don't *merely* leave the worrying about these other things to other experts, as a kind of division of labor; they do not think there is a puzzle about the standards governing these things. The nature of thermometers is assumed to simply provide the standards of thermometers, and so we leave thermometer standards to the experts not just because we have other things to do, but because all there is to determining what a thermometer ought to do is figuring out how thermometers can best do *what they do*.

We agree that thermometers ought to be calibrated and need not be any particular color. So, we think that some ways of judging a thermometer involve evaluation (e.g. the judgment that this thermometer ought to be calibrated), while some ways of judging a thermometer do not (e.g. the judgment that this thermometer is blue). An appealingly simple explanation of the difference is at hand: some judgments of thermometers involve judging them according to the standards of thermometers. Thermometers are an *evaluative kind*, a kind whose members can be better or worse *qua* members of that kind. Certain judgments of individual thermometers are judgments relevant to assessments of their being

better or worse qua members of their kind (e.g. 'This thermometer is properly calibrated.'), and some are not (e.g. 'This thermometer is blue.'). Judgments according to standards that are equally applicable to all thermometers *in virtue of their kind* are judgments according to standards that are norms. Norms, as I understand them, are those standards according to which we can judge excellence and defect. Evaluative kinds, in turn, are those kinds that have standards of goodness, those kinds whose members can be excellent or defective as kind-members. They are, to borrow a phrase of Judy Thomson's, *goodness-fixing kinds*.[1]

We can appeal to *what thermometers are* to understand what it would be for an individual thermometer to be better of worse. The nature of thermometers seems to be characterized by the function of reading and displaying temperature; so a specific thermometer is better or worse *qua* thermometer insofar as it is better or worse at performing this characteristic function. Generalizing, this is a picture on which evaluative kinds differ from non-evaluative kinds in virtue of being *functional* kinds.[2] Individuals can be better or worse members of their kinds in virtue of being better or worse *at serving their function*.

Thermometers are thus evaluable *qua* thermometers in ways that non-thermometers are not—they are evaluable as defective or excellent thermometers. The judgments we can make of a thermometer in virtue of being a thermometer differ from the judgments we can make of non-thermometers. Only the former can be subject to evaluations of defect, for example. A thermometer is defective when it fails to satisfy the norms of thermometers (and thus is unable to serve the function of thermometers—its own function) while a kitchen knife is not defective when it cannot satisfy the norms of thermometers (and is thereby unable to serve the function of thermometers). Being a member of an evaluative kind is a precondition on evaluation according to the norms of that kind.[3]

[1] For Thomson's discussion of goodness-fixing kinds, see her (2008) Ch 1.

[2] This notion of 'functional kinds' is meant to be metaphysically substantial. Contrast this with Millikan's linguistic concept *function categories*, see her (1993), esp. pp 21–22. *Functional categories* is a concept that picks out common nouns and noun phrases that group individuals not by current properties, activities or dispositions, but instead by their function. Her examples are thermometer, can opener, heart, kidney, greeting ritual, mating display, fleeing behavior, stalking behavior, etc. Function categories are introduced as a linguistic type, but my focus here on functional kinds is perhaps best seen as the metaphysical correlate to that linguistic grouping.

[3] While we could truthfully say that the knife would *make* a bad thermometer (or it would be bad *as* a thermometer), we cannot say that it *is* a bad or defective thermometer. It isn't any kind of thermometer, and being a bad thermometer is a *way* of being a thermometer.

Understanding thermometers as a kind with a characteristic function (i.e. as a functional kind) and what it is to be a good thermometer as being a thermometer that performs its function well, is a specific instance of what I call **Functional Goodness**.[4]

Functional Goodness: For any functional kind, K, any individual K-member more suited to serve the function of Ks is a better K-member than one that is not as suited to serve that function.

While any object can be judged according to how well it performs an activity, Functional Goodness picks out how we can evaluate a functionally understood individual in relationship to its characteristic function. Functional Goodness picks out a way of evaluating individuals as better or worse in virtue of being more or less suited to perform a function.

This fits with another natural commitment in the area—**Attributive Goodness**.[5]

Attributive Goodness: Evaluative adjectives, like 'good,' have semantic content primarily as attributive adjectives.

One mark of attributive adjectives is their semantic inseparability from a containing noun-phrase. An adjective is semantically inseparable from a containing noun-phrase when the semantic content is not preserved when the adjective is used alone. For example: 'Mickey is a large mouse,' contains a use of the attributive adjective 'large.' 'Large' is a semantically inseparable part of the adjectival noun-phrase 'large mouse;' while 'Mickey is a large mouse' entails the truth of 'Mickey is a mouse,' it does not entail the truth of 'Mickey is large *simpliciter*.' Adjectives that are not attributive are predicative, and it is a mark of predicative adjectives that they can be semantically separated from their adjectival noun-phrase. 'Mickey is a red mouse' entails both 'Mickey is red *simpliciter*' and 'Mickey is a mouse.'[6] Attributive

[4]Functional Goodness gives an ordinal conception of 'better than' and 'worse than' to which 'good' and 'bad' are relative. I think this is the right way to understand goodness but an argument in support of this position isn't the purpose of this chapter, nor do the main arguments of this chapter turn on this ordinal conception.

[5]Discussions of attributive uses of 'good' seems to begin with Ross (1930) (see especially pp 65–67) and are developed in Geach (1956) and Ziff (1960) and more recently by Thomson (1997) and (2008) (following Geach, see especially chapters 1 and 2) and Finlay (2004) and (forthcoming) and Finlay and Björnsson (2010) (following Ziff's end-relational approach).

[6]'Large' is attributive in a much weaker sense than I am suggesting 'good' is. 'Large' is attributive because of comparative aspects of 'large' and the contrast class is given by the noun, like 'mouse.' It's not just that

Goodness holds that 'good' and other evaluative adjectives are attributive adjectives—they are like 'large' rather than like 'red.'[7] When we judge that this thermometer is good, we must mean that it is good *as something* and the obvious position is that it is good as member of its kind, as a thermometer.[8] Moreover, if functional goodness is believed, it is good *as the thing it is* by being suited to serve its function.

The foregoing account is an instance of *constitutivism*, a popular strategy of grounding the norms in what is constitutive of the objects those norms are used to evaluate. No one troubled by an inability to account for what makes a good thermometer *good* or worries about why a thermometer that has been recently calibrated is better than one that hasn't. No one wonders why my spade isn't defective when it isn't able to give an accurate reading of your temperature. Even if we don't have a ready account of what the norms are, it is unremarkable that there are better and worse thermometer and that it is (all and only) *thermometers* that are uniquely evaluable according to the norms of thermometers.[9] The appeal is obvious—we have a explanation of thermometers as the sorts of objects that can be better or worse at being the kind they are and this provides a ready explanation of why

'good' needs a contrast class, it needs a contrast class with a shared nature to *give* 'good' its specific content. What properties are required for being 'good' changes for the kind, K, that 'good' is attributed of. What is required to be 'large' does not change, only how much of it you need, given the contrast class. So, why think that 'good' is like 'large' at all?

I do think that 'large' and 'good' are disanalogous in this (and other important) ways. But all I need here is that they share this in common: to know what would count as a 'large' or a 'good' k-member, you need to know what kind of thing you are on about. Something cannot just be good simpliciter, and it shares this property with other adjectives like 'large.'

There is another interesting disanalogy here: restrictions on the comparisons you can make with 'better' and 'worse' that don't hold for 'larger' and 'smaller.' For example, it would make sense of my desk, to ask of whether it is large, as far as objects in this room go. But it would make no sense to ask of the desk, whether it is good as far as objects in this room go. This is because (you'll have to trust me on this) there are objects other than desks in this room, and it only makes sense to ask of two things, which is the better one, if they are of the same goodness-fixing kind.

[7]While I claim 'red' isn't an attributive adjective, because 'Mickey is a red mouse' entails both that 'Mickey is a mouse' and that 'Mickey is red,' it seems that not all uses of color adjectives work this way. So, for example, to know of a certain colored object whether it counts as a red example of that object, we might need to know what that object is. The color that counts as red for hair would not count as red for paint, for example. Does this mean that 'red' is an attributive adjective? I think there might be questions worth considering here, but showing that all adjectives were semantically inseparable would not undermine the claim here.

[8]Schroeder has nice discussion of evaluations of goodness of objects that are more than one kind. I address concerns about this form of multiplicity in Ch 3. For Schroeder's discussion, see his (2007).

[9]These strong negative universals are clearly too strong, and perhaps, as has been pointed out to me, even if they are true, they ought not to be. The rhetorical force here is important, however. Those who accept thermometers as unproblematically norm-governed might be led to see the commonality with non-artifactual norm-governed kinds. Those who do not will need to find the commonality independently motivating.

these facts aren't troubling. The case of thermometers is straightforward; thermometers are a functional kind—their natures are explained by their function of taking and displaying the temperature of objects or the environment—and better individual thermometers are those that are better at serving that function.

Constitutivism understood as such is a general metanormative project that explains norms by appeal to the nature of the kind governed by those norms.[10] Constitutivism is characterized by what I call the **Metanormative Commitment**.

Metanormative Commitment: The norms or standards that govern an object, x *qua K*, can be explained in virtue of the constitutive nature of the kind, K, that x is.

According to the Metanormative Commitment, norms governing an individual are explained by appeal to the constitutive nature of the kind that individual is. The norms that govern a particular thermometer govern all thermometers as such, in virtue of the function that characterizes the nature all thermometers share.

So, we can see that there is a general sketch of a constitutivist argument that we've led up to.

Sketchy Metanormative Constitutivism

1. Evaluative kinds, Ks, are such that for each K, K has a constitutive function, Z.
2. *Promissory note for explanation of when an individual x is a member of kind K.*
3. An individual, x, has Z if and only if x is a member of an evaluative kind with that Z. (Bridge Principle)
4. An x that has Z is better to the extent that it is suited to fulfill Z and worse to the extent that it is not.[11] (Functional Goodness)
5. Properties that further the fulfilling of Z are virtues in individuals that have Z.[12] (Good For)

[10] Constitutivists need not explain the nature of evaluative kinds in terms of functional kinds, but this is one natural way we already seem to understand the norms of artifacts and (possibly) biological kinds.

[11] Because an x could be good at Z-ing without ever being a position to Z, we need the claim to be that it is better to the extent that it is capable of doing so, rather than to the extent that it does. This becomes additionally important in Ch 3.

[12] There is a possible worry that not every property that makes an individual better at fulfilling its function is good for it. Suppose, like some implements of war, the purpose of some object, Explody, is its own destruction. Then properties that allowed Explody to self-destruct more effectively would be good for Explody. But this seems counter-intuitive. It isn't good *for* Explody to explode, we might think. It also might not be good for *anything* else that Explody to explode. So, being a good bomb might not be good for Explody.

6. Properties that frustrate the fulfilling of Z are vices in individuals who have Z. (Correlate of Good For)
7. A norm, N, governs an individual, x, when x would be better insofar as x satisfied N and worse insofar as x failed to satisfy N. (Norm Governance)
8. So, norms that govern individual Ks are those standards that require properties that further the fulfilling of the Z of Ks.

This is a general sketch towards the Metanormative Commitment, characterized roughly in the final claim, 8. The sketch promises, but does not give, an explanation of what the conditions are for an individual x to be a K. This is a strength of this particular sketch's sketchiness, because it will allow for differing constitutivist accounts to diverge here. But the argument begins with what is constitutive of evaluative kinds, having a function, and moves to explaining norms that are applied to *individual* kind-members in terms of this constitutive feature. It is the *individual* that is the object of evaluation, but the *evaluative kind* that is understood functionally. So no constitutivist account gets a free pass on the promissory note at 2; the constitutivist must give an account of about what makes any particular individual a member of an evaluative kind and thus a proper object of evaluation according to the norms of that kind.

Putting this task temporarily to the side, the first things that should strike us are how general this account is and how powerful it could be if successful. It aims to provide an account of norms for any evaluative kind.

2.2 GENERAL CONSTITUTIVISM

In leaving open exactly how individuals share in kind-hood and what sort of constituting activity or function characterizes kinds, the constitutivist account above is quite general. On this way of understanding constitutive accounts of norms, then, they aim to ground norms in the nature of evaluative kinds. The most well-known constitutive accounts are those of practical rationality, which aim to show that the norms of practical reason come from (are constitutive of) the nature of agents or agency as such. But though these accounts, notably championed by Korsgaard and Velleman, are the most explicitly developed as constitutive

accounts, the general approach sketched above has a much wider following.

On this understanding of constitutivism even some instrumentalist accounts of practical reason count as constitutivist accounts.[13] This broad conception of constitutivism has support in the existing literature. For example, Railton's 'low-brow' constitutivist accounts hold that it is constitutive of agents that they aim at what they want; agents, as such, are disposed to follow the instrumental principle.[14] 'High-brow' accounts on the same view are those that pick out some particular end that agents as such must aim at. According to one particular high-brow view of agency, action necessarily aims at 'the Good.'[15] Railton's accounts closely link what is constitutive of agents as such and what the conditions are for any individual being an agent. His low- and high-brow accounts take various aims or activities to be characteristic of the kind 'agent' and the norms of agency to be explainable in terms of those aims or activities.

The low- and high-brow views of agency are both commitments about the nature of agency, and have correspondingly low- and high-brow forms of constitutivism. These views are constitutive, in our broad sense, because they take the nature of kinds to ground the norms by which individual members of those kinds are governed. On these views, what it is to be an agent is to aim at what one wants or what one finds to be good, so an action is necessarily assessable by standards that come from this constitutive aim. So, the nature of agency on these views accounts for the norms of agents without making them dependent on contingent facts about individual agents or from being imposed on agents illegitimately.

On these views, being good *qua* agent is to be good according to the standards of agents. Constitutivism of the sorts given by low- and high-brow views aims to provide an account of practical rationality. It takes practical rationality to be a matter of excellence or goodness according to the norms that govern agents *qua* agents, and takes agents as such to be characterized by certain aims. The norms governing agents will then be those that further the satisfaction of those aims, and an agent will be better or worse *qua* agent, which is to

[13] Dreier and Williams, for example, provide instrumentalist views with constitutive explanatory accounts that fit this account characterized by the Metanormative Commitment. See Dreier (1997) and (2001) for examples.

[14] For a standard introduction to this broad view of constitutivist projects, see Railton (1997) pp 62-69. For an instance of low-brow instrumental constitutivism, see Dreier (2001)

[15] "...in choosing an action we place it (or find it to be) in a positive evaluative light, and deem it choice-worthy." Railton (1997), p 62

say *more practically rational* insofar as she satisfies the norms that further the satisfaction of her essential aims, those she must have as an agent.

The most well-known constitutivist accounts in the literature are accounts of practical rationality.[16] Importantly, though, constitutivism is not specific to any *particular* evaluative kind. The general characterization of the account here generalizes to all evaluative kinds. So, if it is successful, constitutive accounts should allow us to account for norms governing all evaluative kinds.[17]

Constitutivism is a general way of explaining for certain Ks (those that admit of goodness) what norms any individual K-member is held to, by appealing to what is constitutive of Ks; the Metanormative Commitment is the essential feature of this move—it holds that norms that govern some individual K-member do so in virtue of the nature of Ks. By supplying an account of the nature of evaluable kinds, either by Functional Goodness or some other account, norms are specifiable as those standards whose satisfaction makes an individual better at achieving something essential for its kind. Taking evaluative kinds to be functional kinds, I claimed above, accounts for how evaluative kinds—as kinds with essential functions—could have internal standards that individual members could be better or worse at satisfying and that govern all individual kind-members *as such*. But we still don't have an account of the conditions for an individual to be a member of an evaluative functional kind, which is an essential part of the constitutivist story.

Constitutivist accounts appeal to an intuitive distinction between kinds that can have internal standards of goodness and those that cannot. Kinds that do not have internal standards of goodness are those for which the only way to be good is to be good is *for*

[16] Those accounts accept that agency or action is defined as having a necessary function or aim and then appeal to this function to explain how good action is action that is well-situated to further, or which does further, the function of action/agency. This is, essentially, what Setiya calls "ethical rationalism"—it aims to explain the standards of practical reason as coming from the very nature of rationality itself. He does not characterize ethical rationalism as a type of constitutivism, but instead takes it to be a more descriptive term for the same commitment. I think this is because his focus is only on the practical. I think he would be amenable to ethical rationalism in his sense being a type of constitutivism in the broad sense laid out above. For Setiya's presentation of ethical rationalism and extended argument against it, see his (2007).

[17] The following recognized constitutivists will hopefully all fall under the definition I'm giving here: Korsgaard (1996) and (2009); Velleman (2000); Rosati (2003); Ferrero (2009). It will also include some who are not commonly thought of as constitutivists, but I think ought to count: e.g. Dreier (2001), Bratman (2009), also, I think the argument in ch 3 of Enoch (2011) would count as constitutivist on this reading, though these claims are controversial, and I do not defend them here.

something, not *as the thing they are*. For example, a sample of the chemical element gold cannot be better or worse *qua gold* than any other sample of gold (despite possibly being very different in, say, size or shape, from other samples of gold). A sample of gold is better *at conducting electricity* than a sample of lead and worse than a sample of silver, and one sample of iron might be better *for holding open the door* than another, because, say, the first weighs 20kg but the second only 20g. But no chemical elements admit of internal standards of goodness. Constitutive accounts that appeal to Functional Goodness will be able to exclude such elemental kinds from the set of evaluative kinds by recognizing that gold is not a functional kind.[18]

The constitutivist project grounds norms in the nature of the object of the norms in question. Constitutive facts are an obvious place to look for such a source, but not all constitutive aspects of kinds are going to themselves be a source of norms (material constitution, for example, won't help). The task of the constitutivist, then, will be to account for how evaluative kinds have their constitutive functions (or, more generally, their constitutive natures). Constitutivist approaches divide kinds governed by internal standards from those that are not—and in doing so, they aim to explain the bindingness of normative standards by revealing them to be grounded in the nature of the items they govern. For those worried about what makes standards genuine and legitimately applicable to the individuals they apply to, constitutivism has thus proven an attractive strategy.[19]

This requires both an account of the constitutive nature of kinds and an account of how individuals share in kind-membership of the constitutive kind. If both these projects are successful then it would open the way for an account showing how, say, we could derive the standards of practical reason from the nature of agency. The procedure, also, must be generalizable, providing an account for all legitimate norms (and perhaps a way of determining the legitimate from the illegitimate-yet-appealing norms).[20]

[18] They do, however, have the challenging burden of showing that all kinds we take to be evaluative can be understood as characterized by a constitutive function. This burden will be one of our focuses in the next chapter.

[19] This has proven attractive because of the otherwise seemingly intractable conflict between internalists and externalists about reasons. Velleman, for example, motivates his constitutivist project by the desire to find internal reasons that do not leave open the question of motivation that would make them relativized. For this discussion, see Velleman (1996) esp. pp 179-80, where he claims to be seeking the benefits of externalism and internalism without their burdens.

[20] This "must" might come off as too strong. I think it's a desideratum for a constitutive account that it

This kind of approach is attractive because the norms it supplies seem neither arbitrary nor mysterious. Locating the origin of norms in the objects of evaluation, without appeal to our own interests, quiets the fear of arbitrariness. Because the nature of an individual's kind grounds the norms that govern that individual, norms are not arbitrary; norms apply equally to all individuals that are members of the kind that grounds the norms in question. Accounting for this grounding in the functional nature of the evaluative kind makes the legitimacy of the norms' application non-mysterious.[21]

Moreover, the focus on Attributive Goodness, the claim that what it is to be good or bad is a matter of being a good or bad *qua* kind-member, makes the constitutivist project particularly resistant to charges of 'funny business.'[22,23] On this account, goodness and badness do not pick out properties that inhere arbitrarily in otherwise valueless objects. You cannot say that apples are good *simpliciter*, predicatively speaking. This apple might be a good one (that is, a good *apple*), but this does not imply that there is some property of goodness in the apple that is shared by good hearts, good reasons, and good thermometers. If successful, the constitutivist explains the standards binding kind-members without making norms merely our own constructs, dependent on our own interests and desires, and also does not commit us to finding some property called *goodness* that is somehow instantiated in otherwise valueless objects.

Constitutivism purports to give us no stronger nor weaker a theory than we want—we are to end up with an account that is maximally satisfying and minimally troubling. If successful, norms are explained as governing individuals as a consequence of what those individuals are. Constitutivism is a general way of explaining for certain Ks (those that

be generalizable because it seems that our demands for legitimacy of norms requires a unified account and that without good reason, for which I can see no hope of expecting provision, the failure to generalize a constitutivist account of practical reason to the norms of natural biological kinds and artifacts, at least, will tend to support a sort of fictionalism of norms of these types that I find unacceptable. But your intuitions and their mileage might vary.

[21]Though, there is the risk of functions or functional kinds themselves becoming mysterious, which is an issue I address in other work.

[22]I prefer Belnap's 'funniness' terminology to Mackie's 'queerness,' though I take them to pick out the same concerning difference. If you prefer, you may replace 'funny' with 'queer' and not lose anything of interest.

[23]The history of Attributive Goodness is anti-funny-business oriented. Though, there is a worry, that this advantage of making evaluative predicates less funny is gained at the expense of making evaluative kinds more funny.

admit of goodness) what norms any individual K-member is held to, by appealing to what Ks are; the Metanormative Commitment is an endorsement of this move.

2.3 CONSTITUTIVISM'S CRITICS

Unfortunately, not everyone finds constitutivism minimally troubling. Despite its appealing pitch, constitutivism seems to be dogged by doubts which seem to strike both advocates and opponents of the constitutivist picture. These doubts are voiced most pointedly by those who frame their concerns around the so-called "Problem of Bad Action" and the "Violability Worry."[24]

Despite assurances to the contrary, critics of constitutivism worry that we cannot have it all; we cannot explain norms by appeal to constitutive facts about kinds. Though the goal of constitutivism is to give an account of norms by appealing to the natures of the kinds they govern, the worry is that the account is unsatisfactory because norms that come from the constitutive nature of the objects they evaluate cannot be violated by those objects. Briefly, constitutivists are thought to be unable to provide norms that are violable and be unable to account for defective or bad kind-members.[25]

Constitutivists need to account for the fact that norms are used to evaluate *individuals*, but it is *kinds* that are primarily understood as having functions. To account for this, constitutivists appeal to the central role they take norms to play in making individuals members of norm-governed kinds; the norms are taken both to account for individuals' kind-membership and evaluate those individuals as members of that kind. This dual role for norms gives rise to the two worries that are thought to follow as consequences from this move. Both kinds and norms become the objects of concern on this picture—kinds become the focus of the Problem of Bad Action and norms become the focus of the Violability Worry.

[24]So-called because most constitutivist accounts are pitched as accounts of practical reason, and so the most common object of evaluation in constitutivist theories are actions (or, confusingly both actions and agents).

[25]For worries of this nature by critics, see Railton (1997), Clark (2001), Barandalla and Ridge (2011), Lavin (2004). For worries voiced by constitutivists, see: Velleman's Introduction to (2000) and Korsgaard (2009), esp. Ch 8.

One way to understand the two problems is as separate challenges for accounting for two phenomena: bad individuals and violable norms. Understood this way, the Problem of Bad Action is a worry about whether the constitutivist project can account for the actual phenomena of individuals that are negatively assessable according to norms of their kinds, i.e. bad actions. If it cannot, then it seems to entail the impossibility of actually existing phenomena. As it turns out, we live in a world almost exclusively populated by (more or less) negatively assessable individuals; our world may be the best of all possible worlds, but it seems to be full of things that are not the best of all possible things.

The Violability Worry can thus be understood as a correlative worry about the status of norms that the constitutivist provides. Norms on the constitutivist picture are evaluative of the individual members of kinds those norms constitute. If norms serving this purpose rule out defect, they also will be necessarily satisfied by individual kind members. In such a case, norms would be inviolable; they would be satisfied necessarily by everything they applied to. So, the worry is that the norms the constitutivist picture provides are not capable of being violated by the individuals to which they apply.

However, the two worries seem to extend beyond accounting for these phenomena. Even if it were the case that all individuals were maximally assessable according to the norms that governed them, the worries would not be quieted. The worries aren't about accounting for phenomena, they are about the possibility of evaluation. The Problem of Bad Action would be a problem even if all individuals maximally satisfied all norms, even if, as it turned out, all actions happened to be perfect ones. The Problem of Bad Action doesn't just entail that no individuals are bad, it entails no individuals *could be* bad. It's a claim about the very possibility of bad actions. Understood as such, it comes with the Violability Worry. The possibility of negatively assessable individuals and the possibility of norms that can be violated are two sides of the same coin for constitutivists. The Violability Worry threatens that the standards given by constitutive accounts are the sort of standards that cannot possibly be violated by individuals to which they apply. For a constitutivist this just entails that no individuals could be negatively assessable, because it is only by being in violation of a norm that an individual could be negatively assessable.[26]

[26] And note, importantly, that this understanding already appeals to the distinction between violating and

This joint understanding of the Problem of Bad Action and the Violability Worry is that constitutivists define away the very possibility of both defect and violable norms. All norms that can be derived from the nature of evaluative kinds will be necessarily met by all individuals to which they apply. In the rest of the chapter, I will refer to both the Problem of Bad Action and the Violability Worry as simply the Violability Worry, as they are inseparable on the constitutivist account and the latter involves more general terminology.

The Violability Worry is a concern because if the only norms that can be derived from constitutivist accounts are met by all individuals to which they apply, constitutivism is a dead end as a way of explaining norms. It is essential to our understanding of norms that they are necessarily capable of being violated and individual members of evaluative kinds must be capable of being negatively evaluated according to those norms. Otherwise, the constitutivist project fails.[27]

The force of the joint concern of the Violability Worry seems to be felt even by constitutivists themselves and resists even seemingly successful responses. Neither critics nor constitutivists themselves seem satisfied by responses in print. Constitutivists, in particular, seem pre-occupied with their ability to account for violable norms and bad individual kind-members.[28] It is particularly interesting that this worry seems so resistant to replies. I hope to show that despite the persistence of the worries, they can be answered. And moreover, the answer can help explain why the worries were so persistent in the first place. I will argue that we've misunderstood the consequences of the way current constitutivists relate individuals and their kinds, but those troubled by the violability worry are looking in the right place for the weak aspect of constitutivism. There is, I will argue, a genuine problem in the vicinity, but the violability worry has distracted us from the real issue.

---

not satisfying. The worry is not that the norms are met by *everything at all*, but that they are met by everything that they *apply to*. So, there's already an implicit distinction being made here to legitimate and illegitimate norm application.

[27] It is worth noting that this is not the only strain of objection to the constitutivist picture. In addition to worries that constitutive norms cannot be appropriately violable and thus evaluative, some also worry that constitutive norms lack the normative or binding categorical force that (at least practical) norms are thought to require. This objection is beyond the scope of this project, but any full defense of a constitutivist project owes these critics a compelling response. It is the subject of further work. For these criticisms, see Enoch (2006), Shah (2003), Mele (2004), Hussain (2004), and Tiffany (2012). Velleman responds in (2004) pp 291-97.

[28] See, especially, Korsgaard's Ch 8 of her (2009) and Velleman's Introduction to (2000) for examples of this pre-occupation.

First, we'll look at some of the more common expressions of the general Violability Worry in the literature, given both by critics and constitutivists themselves. I'll then give a natural response to these arguments on behalf of both Korsgaard and Velleman and argue that it generalizes and gives a decisive rebuttal to the Violability Worry. Despite all this, the nagging sensation that something is amiss remains. So, in the following section I will try to motivate a worry that persists even if we remain fully satisfied that the positions of Korsgaard and Velleman are responsive to the Violability Worry. I will then argue that this new worry raises a challenge that current constitutivists cannot meet with the commitments they have.

From here, we will see if there is another way for a constitutivist program to maintain the strategy of explaining norms as derivable from the constitutive nature of what they govern that is able to explain this connection. We will retain the appealing properties of constitutivism that recommend it to begin with, and we will be able to, I argue, successfully allay our concerns about violability for good. In the next chapter, the promissory notes given at the end of this section will be cashed out.

2.4 EVERYONE'S A CRITIC

The Violability Worry seems based in the thought that constitutivism reduces norms to the status of inviolable alethic facts.[29] Consider the fairly explicit appeal to such a claim in what Railton calls the "danger of pulling the claws of criticism":

> Assume, for example, that the connection between taking oneself to have an end and according that end deliberative weight is a substantively necessary, non-analytic connection of the same modality as the connection between being gold and having atomic number 79. What would we then be able to say by way of *criticism* of an agent who refused to give deliberative weight to his own acknowledged end E? Would he be 'necessarily deliberatively defective' or perhaps 'self-defeatingly irrational'?
>
> If the constitutive argument is right, we cannot even raise the question! To fail to take oneself as according E deliberative weight is to fail to acknowledge E as an end. But then the agent cannot be 'self-defeating' or 'irrational' with respect to E. An analogy: to discover

[29] Though there are multiple types of alethic modals—metaphysical and logical necessity are the most common—I think putting this distinction in terms of alethic modality is fine here.

> that he metal in the sample tray on one's laboratory bench has atomic number 82 is not to discover that it is 'defective gold,' but rather that it is not gold at all.
>
> A similar problem confronts all constitutive arguments.[30]

If norms become incapable of being used for criticism on the constitutivist account, this amounts to 'pulling the claws of criticism.' Norms on Railton's view of the constitutivist project are standards of kind-membership, and so instead of being in a position criticize an action when it is contrary to reason (and thus in violation of norms), you simply discover that the behavior or act is not an action at all. If all possible Ks meet the norms of Ks, then norms become mere sorting standards, because they are necessarily satisfied by the individuals they seem to exist to evaluate. The concern arises because of how constitutivist accounts are taken to closely link the meeting of norms with being the sort of individual the norm governs. Korsgaard advocates the constitutivist view as being able to "respond to skeptical worries with ease," but we might see the Violability Worry as a voicing of the thought that it might be a bit *too easy*.[31]

Clark makes a similar objection to Velleman's close connection between the aim of action and rational criticism when he writes:

> Velleman's view makes it impossible to criticize any fully intentional action as being contrary to the weight of reasons... A constitutive goal, for Velleman, is something at which one must aim in order to count as doing the thing in question, in this case believing: "I cannot believe something without accepting it seriously—in an attempt, by me or my cognitive faculties, to arrive at acceptance of the truth"...
>
> The constitutive aim [of action] is autonomy, but as we've just seen, it is a premise of the argument that every fully intentional action is autonomous. So unlike belief, which however full-blown, can and often does fall short of its constitutive aim, fully intentional action never exhibits the practical analogue of falsehood.
>
> This makes it difficult to see how a fully intentional action could ever be rationally criticizable, on Velleman's view. One thing that seems to happen on a regular basis is this: someone does a fully intentional action, but the weight of reasons is against it.[32]

[30] Railton (1997) p 70.

[31] Korsgaard (2009).

[32] Clark (2001) pp 581–82, Clark quotes Velleman (2000) p 710.

Note here that Clark's objection seems to come across as a criticism that Velleman cannot account for the existing phenomena when he writes "one thing that seems to happen on a regular basis is this..."

Clark is worried that for this view proposed by Velleman, where the nature of action is having an aim, if satisfying the aim of action is a condition on being an action, then any fully intentional action is fully autonomous, and thus, incapable of rational criticism. This is an instance of a more general concern about the role of standards—if what it is to be a K is to meet certain standards, then those standards cannot also serve as a way to differentiate *good* kind-members from *bad* kind-members. Constitutive standards can't seem to serve both as necessary conditions on K-hood *and* as evaluative standards of the same kind.

This worry isn't lost on Velleman or Korsgaard. In developing the role of a specific kind of constitutive standard, which she calls a constitutive principle, Korsgaard recognizes this threat. She explains the problem succinctly:

> [If] it is the essential nature of [a kind] that it have a certain metaphysical property... but in order to have that metaphysical property it must have a certain normative property... then this explains why the [individual] must meet the normative standard: *it just isn't [a member of the kind]* if it doesn't. But it also seems as if it explains it rather too well, for it seem to imply that only good [individuals] really [are kind-members], and that there is nothing left for bad [kind-members] to be.[33]

Velleman also recognizes the problem Clark raises, going so far as having altered his account in response to Clark's challenge. Velleman writes that "If autonomy were the constitutive aim of action, then every instance of action... would turn out to be a success."[34] Velleman acknowledges that it can't be the aim of every action that it be autonomous, because autonomy is a precondition of action. If this were the case it would mean that something only was under a standard or principle if it already met it.

The Violability Worry seems to strike everyone as a particularly *concerning* issue for constitutivist accounts of practical reason (and norms in general)—many take the worry to render the constitutivist project hopeless. If constitutivist accounts can't provide genuine distance between kind-membership and perfection, then these worries are devastating. All involved, whether advocates for or adversaries of constitutivism, seem to think the problem is a real threat to constitutivism.

[33]Korsgaard (2009) p 160; I've replaced uses of 'action' with more general 'individuals' and 'kind-members,' and left Korsgaard's own italics.

[34]Velleman's Introduction to (2000) n 37.

2.5 DISSOLVING THE VIOLABILITY WORRY

Though the Violability Worry leads to a mystery, the mystery is not how it could possibly be resolved but *why it strikes us as so difficult to resolve.* The worry takes the form of a conditional: If satisfying norm N is a constitutive condition on being an individual K-member, then N is necessarily met by every individual K-member, thus making every individual K-member a perfect K-member according to N *and* N inviolable. The so-called problem strikes many as very worrying, but it is a conditional worry where the antecedent is taken to be a commitment of constitutivists. But, as constitutivists themselves have been at pains to point out, this is a mistake.

On a very strong reading of 'satisfies norm N' where this means 'maximally' or 'fully satisfies,' this conditional is true. But importantly, such a reading of the antecedent is incredibly strong, strong enough that it ought to be incumbent on critics to show that constitutivists are committed to such a position. In fact, no constitutivist I am aware of is committed to such a strong connection between the satisfaction of norms and membership in evaluative kinds. As a result, no constitutivist is, for this reason, committed to the consequent of our Violability Worry conditional above, for any norm. The mystery here is why people continue to be worried about constitutivism on this account.

To see why a very strong connection between individuals and kinds is needed, consider the Violability Worry again. One main consequence of the Violability Worry is that the norms that constitutivist accounts provide are inviolable, that is, they do not satisfy **Violability**.

Violability : That a norm N applies to x and requires p does not logically or metaphysically imply that x satisfies p.

Violable norms are essential to the possibility of defect. On the constitutivist story, N applies to x when x is a member of kind K and N is derivable from the constitutive nature of Ks. So, Violability is a requirement that it be possible for something to both be a member of kind K and to fail to satisfy a constitutive norm of Ks. This means that if the Violability Worry is really a problem, it must be because the constitutivist has made it a condition on being a member of kind, K, to satisfy every norm of Ks in such a way that prevents negative

assessment according to those norms. This is an extreme way to link individuals and kinds.

As it turns out, there are kinds that we understand as constituted by standards in exactly this way (e.g. elemental natural kinds and geometrical objects). These kinds are those we classified earlier as non-evaluative; their constitutive standards give conditions for kind-hood, but no possibility of being better or worse *qua* their kind. So, this extreme constitutive understanding of kinds, rather surprisingly, entails the impossibility of *evaluative* kinds.

This is anticipated in the remarks of Railton, quoted above, where he worries that constitutivist arguments make the relationship between an object and what is constitutively required of it like the relationship that holds between gold and the atomic number 79. We can see the desired contrast between norms and alethic requirements starkly by considering the laws of natural science or real definitions in geometry. Consider the real definition of the geometrical object the circle:

Circle$_D$: A circle is a line composed of points equidistant from a center point.

Circle$_D$ gives us all the information needed about a two dimensional geometric figure, the circle, to know how to construct one and understand its relationship to its constituents. It tells us that a circle is constituted by a two dimensional figure, a line, when the constituents of that line, its points, all have the same relational property of being equidistant from some center point. Circle$_D$, however, does not give us a norm of circles, because it is a requirement on norms that they be violable, and Circle$_D$ does not meet the conditions given in Violability. The constituting conditions Circle$_D$ gives are necessarily met by all objects to which it applies. All circles satisfy Circle$_D$ and they all satisfy it equally well.[35]

The Violability Worry amounts to a challenge that the norms generated by constitutivist accounts are like alethic requirements (i.e. of the type of Circle$_D$). If constitutivist accounts take constitutive norms to be like alethic requirements, fully satisfied as a condition of individuals being a member of the constituted kind, then the Violability Worry is inescapable. There are no circles that are negatively assessable according to Circle$_D$. Circle$_D$ is, in

[35]There are, of course, non-geometrical circles that do not adhere to Circle$_D$—e.g., ones that children draw to represent balloons. But circles, the geometrical objects that are the object of study in geometry, all do satisfy Circle$_D$ and have meeting Circle$_D$ as their only mark.

principle, *inviolable*.[36] All circles are equal, as far as the standards of circles go. In addition to there being no *bad* circles, there are also no *good* circles. If constitutivists make the constituting standards of evaluative kinds like real definitions, those standards would be inviolable and the very possibility of evaluation would be lost. The worry is thus extreme: constitutivism might rule out the very possibility of evaluation.

It's pretty clear that critics think constitutivism has a problem accounting for how individuals and kinds are related in a way that can make norms violable and some individuals negatively assessable. What is not clear is why critics think constitutivists account for the relationship between individuals and kinds on the model of alethic constitutive standards like Circle_D.

The short answer to the Violability Worry is that constitutivists do not take constitutive norms to be like alethic standards. There are two ways that current constitutivists explain the constitutive connection between kinds, individuals, and norms that does not take the extreme position of making constitutive norms like alethic standards but allows them to still account for the relationship between individuals and kinds. One explains the relevant condition on having the function as minimally meeting scalar norms that come from the function.[37,38] The other understands the relevant condition on having the function as having the disposition to perform the function.[39]

Both the scalar and dispositional accounts allow for an individual x to have the characteristic function or goal Z, and thus be governed by the norms derived from Z, without ruling out, for any particular norm N, that x is negatively assessable according to N. The scalar account allows that it might be negatively assessable according to N, while still minimally

[36]There are many things that are not composed of a line whose points are all equidistant from a center point, but none of those things are geometrical circles. Circle_D is thus inviolable because it gives a necessary condition on kind-hood. Further, *what it is* to be a circle is to be composed of a line whose points are equidistant from a center point; there is nothing that could be a circle that failed according to this condition. So, the condition given in Circle_D is inviolable and there is no possibility of bad circles.

[37]Scalar norms are those that require satisfaction of standards that require having some property that can be had to different degrees (e.g. sharpness or buoyancy.)

[38]Though scalar norms are more common, I think, it's also possible to have a set of binary norms that function as I describe scalar norms here; such a set would require meeting some minimum number of the binary norms for kind-hood and for each additional member of the set satisfied, the individual in question would be better than if it hadn't satisfied it.

[39]In the penultimate section of this chapter I argue that the metaphorical "aiming" talk that some constitutivists use to explain the constitutive nature of evaluative kinds and their norms must be dispositionally understood.

satisfying N (e.g. x is not *as sharp* as it could be, but is still minimally sharp). The dispositional account would allow that x might be negatively assessable in virtue of not satisfying N, while still being disposed to satisfy N.

Neither account makes having a characteristic function a matter of being maximally good according to any particular norm. Constitutivists can account for negatively evaluable kinds (and thus evaluable kinds) and violable norms (and thus evaluation) because they account for the relationship between individual members of kinds and the kind in question that does not entail that every kind-member excels according to the norms of the kind as a condition of kind-membership. Those worried about Violability seem to think that there is no connection loose enough to both account for negatively assessable yet legitimate instances of kinds, but this is not so.

Constitutivists can also show how norms on their account are essentially violable. They can show that for any norm N that is constitutive of a K, that N is constitutive of Ks does not logically entail that any member of K satisfies N. Constitutivists can account for norms that meet Violability because they do not, contra the worries of all involved, for any individual norm, make *what it is to be a K-member* satisfying that norm.

Both these accounts endorse versions of what I call the **Threshold Commitment**.

Threshold Commitment: An individual or object must at least partially satisfy the constitutive requirements of a kind, or partially meet the constitutive aim of a kind, in order to be a member of that kind.[40]

The Threshold Commitment is the position that it is a necessary condition on some individual, x, being a member of evaluative kind K that x meet some minimal subset of the norms that govern K-members. The commitment identifies having a function with minimally performing that function or being disposed to perform that function and then performing that function better with being better as a kind-member with that constitutive function. Being disposed to satisfy norm N might not immediately seem to require minimally performing

[40]This may seem ambiguous between an intentional and an extensional reading: the former would hold that is was a necessary condition of being a K-member that one satisfy some minimal norm of Ks and the latter would hold that for all K-members, they satisfy some minimal norm of Ks. The former entails the latter, but not vice versa. I intend the stronger reading.

that function, but unless we admit of the possibility of intrinsic masking, all cases of masked dispositions won't be cases of failure to satisfy the norm, and so to be disposed to satisfy the norm N requires that in at least some cases, the norm is satisfied. So, being disposed to satisfy norm N entails minimally being able to perform the function. This is the subject of the penultimate section of this chapter.

The constitutive norms, then, still determine whether or not any individual is a member of the relevant evaluative kind. But the constitutive norms in question admit of gradation; they are the sorts of things that make an individual better, as the thing it is, the better it is at satisfying the norms. So, while it is a binary matter whether an atom has 79 protons, it is not (merely) a binary matter whether a knife is sharp. So, sharpness might seem suited to serve both as a criterion of kind-hood for knives and also a norm of knives. At least, this is the thought behind the Threshold Commitment.

The Threshold Commitment also makes sense of some of our linguistic practices when it comes to attributing (or denying) kind-membership to objects of evaluation.[41] We often make claims like (pointing to the liquid in a mug) "This coffee is so bad, it's not even coffee!"[42] So, if you think that meeting a minimum standard of, say, coffee, is a condition on being an instance of coffee, then rather than being a mere expression of disapproval at the state of the coffee, we could take the utterance as a literally true expression of the metaphysical status of the object in question. Being bad enough at being coffee might lead to no longer *being* coffee at all.[43] Minimal kind-goodness is thus a condition on kind-membership.

Constitutivists have a way to relate individuals to kinds via norm satisfaction that does not entail the bad consequences that seem to so worry critics of constitutivism. By embracing the Threshold Commitment the constitutivist is able to make norm satisfaction a necessary

[41] Of course, it need not make sense of *all* of our linguistic practices. We (dedicated Aristotelians excepted) are generally comfortable referring to blind eyes as eyes, for example. The point here is simply to show that the Threshold Commitment is not without support, not to give an argument in its support via our linguistic practices.

[42] True story: this was said to me by Luca Ferrero. The coffee was quite bad and the conversation that followed included my first formulation of the argument of this chapter. Thanks, Luca!

[43] This is a form of the ancient position that reality is a shadow of perfection, best seen (as far as I know) in Plato's Theory of the Forms. It's a commitment to the thought that the real must have involvement in the perfect. Though it's no longer fashionable to endorse the theory of the forms in most circles, its pull in circumstances of evaluation is generally under-appreciated. For an explicit endorsement by Korsgaard see (2009) p 31.

criterion of kind-membership but avoid the threat of the Violability Worry; norms can serve as the threshold as minimum criteria for kind-membership, while still being violable and allowing for better and worse instances of kinds.

We would be forgiven for assuming that the worries about violability would just go away when the critics recognize the ways that the constitutivist relates individuals to kinds via norm satisfaction à la the Threshold Commitment, but this seems unlikely. It's unlikely because it's so implausible that critics fail to notice that constitutivists *already* have made it explicit that they take minimal norm satisfaction to be a requirement on having a constitutive function. In fact, it's a key part of current constitutivist accounts that norms play this role. It is what gives practical norms their inescapable normativity according to constitutivists that if you fail to accord with them enough you cease being an agent at all. Korsgaard is explicit in *Self-Constitution* that it is only in according minimally with the norms of houses that what would otherwise be a mess of boards, insulation, and roofing shingles is a house at all.

It seems that the most natural reading of current constitutivists like Korsgaard and Velleman is not one where they take constituting norms to be standards like those that constitute gold or circles. But it is only by understanding them in this way that we wind up with inviolable norms or an impossibility of better or worse kind members. Moreover, it seems like we already have a pretty clear understanding of current constitutivist projects that do not entail the Violability Worry. So we now find ourself left with mystery we started with, which is why are people so worried about the Violability Worry. Why then are we still worried?

2.6 THE PROBLEM WITH THRESHOLDS

Despite the seemingly responsive answer to the Violability Worry, no one seems satisfied. The remaining mystery is this: though better or worse instances of kinds and violable norms are possible on current constitutivist pictures, this fact doesn't seem to ease worries about constitutivism's viability. Critics still are unconvinced of the ability to explain norms by

appeal to constitutive natures and the worries continue, despite our best efforts to answer them.

I think that the worries haven't gone away despite available responses to the current arguments about the Violability Worry because there *is* a real problem with the way the constitutivist links kinds and individuals. On any constitutivist account, any particular instance of a kind can be negatively assessed according to any individual norm that applies to it, but the persistence of the worries indicates another problem: there is a related unacceptable consequence.

The Threshold Commitment entails that fully defective kind-members are impossible.[44] Objects that are fully defective according to the norms of a kind cannot be members of that kind, and so no individual kind-member could be fully defective. This is not particularly surprising, as it is what we might expect given some of the motivation for the Threshold Commitment—after all, defect was supposed to be the road to non-being. But this aspect of the Threshold Commitment has consequences that are concerning: it is incompatible with a main use of norms as standards of evaluation.

Violability was one constraint on norms; it held that a norms must be violable, that no standard whose legitimate application entailed its satisfaction could be a norm. Evaluablitiy is also a constraint on norms; in order for a standard to be a norm, the individual the standard governs must be evaluable, judged to be better or worse, according to that standard, not merely be judgable as meeting the standard to some degree or other. Norms are standards that are used to evaluate individuals according to whether the individual assessed is *better* or *worse*; they are not merely sorting standards. This is an essential aspect of the evaluative nature of norms, which I call **Norm Governance**.

Norm Governance: A norm, N, governs an individual, x, when x would be better insofar as x satisfied N and worse insofar as x failed to satisfy N.

[44] 'Fully defective' is a semi-technical term here. It is meant to pick out the state of an individual K-member that is in violation of every norm that governs K to the greatest extent possible for that norm. It is not meant to pick out the most defect possible before an individual ceases being a kind-member. The constitutivist who endorses the Threshold Commitment is committed to the position that there are no kind-members that are fully defective in the first sense, my sense, not the second. All kind-members of evaluative kinds are at least minimally positively assessable as a consequence of the Threshold Commitment.

The problem comes with the comparative conception involved in 'better than.' X being better in virtue of satisfying norm N logically implies that x would be worse if x did not satisfy N. It is instances of the application of this inference that the Threshold Commitment rules out—there are some cases, those at a threshold, where the relevant individual at the threshold would *not be worse* if it failed to satisfy a norm that it in fact does satisfy.

According to the Threshold Commitment, norms serve a special function at the cusp of kind-hood that effectively prevents this consequence. When the norm in question is satisfied, the individual is a member of the kind governed by that norm and thus ought to be better in virtue of satisfying the norm. But this is ruled out because being a better K-member implies it would be a worse K-member if it did not satisfy the norm, and that is not the case. If the individual did not satisfy the norm, it would no longer be a K-member at all: the norm would no longer apply. So, the role the norm plays at the threshold prevents it being fully evaluative; were the norm to cease being satisfied, the object would not be worse in virtue of failing to satisfy the norm. So when the norm *is* satisfied, we cannot say that the object is better in virtue of satisfying it.

The Threshold Commitment entails the impossibility of fully defective kind-members and thus conflicts with Norm Governance. It is unsurprising that the Threshold Commitment makes fully defective kind-members impossible; according to those who endorse the Threshold Commitment, this consequence is a feature of the account, not a bug. But unfortunately, the impossibility of fully defective kind-members severs the obvious and essential connection between norms and evaluations of better and worse. If Norm Governance is true, then any K-member must be capable of being negatively assessed by any norm of Ks. This just entails the possibility of fully defective kind members. The Threshold Commitment makes some applications of norms non-evaluative and rules out some individual members of Ks illicitly, which tracks problems of norm violability and bad kinds. These problems explain why the violability worries were so much more worrying than they warrant.

2.7 A STORY OF TWO KNIVES

Current constitutivists explain how individuals share in kind-membership by appeal to the Threshold Commitment; they take partially satisfying the norms that govern a kind a necessary condition on being a member of that kind. This fills in the promissory note to link individuals that are governed by norms with the kinds that have the norms constitutively. Unfortunately, this way of linking individuals and kinds does not yield a satisfactory accounting of either the possible ways of failing to satisfy norms or the evaluative nature of norms. The Threshold Commitment rules out fully defective individuals and, though constitutivists seem happy with this consequence, the impossibility of full defect leads to an unacceptable conflict with Norm Governance.

So, while we can make sense of bad actions and bad knives on the current constitutivist picture, we cannot account for an individual K-member that fails according to *all* norms of K. An example will illustrate why this is problematic. Consider an object we'll call 'Pointy,' that is a knife in virtue of minimally meeting the norms of knives. That is, 'Pointy' is a knife at the threshold of knife-hood. Were Pointy to fail to meet the last knife norm, Pointy would cease to be a knife (it would fall off the ontological cliff, so to speak). As a result of this failure, Pointy would no longer be evaluable by the norms of knives.

The problem Pointy faces isn't one covered in the Violability Worry; Pointy is a bad knife and is violating many knife norms. The problem Pointy faces is a more perilous one: Pointy is minimally good, in virtue of meeting one last knife norm, which we'll call N^K, but *would not be worse* if it ceased meeting N^K. Consider a nearly identical object, Dull, that is like Pointy in every way, except that Dull fails to minimally satisfy N^K. According to the Threshold Commitment, Dull is not a knife, and so cannot be a worse knife than Pointy. But, as Pointy *is* a knife, and N^K is a knife norm, Pointy ought to be better, in virtue of satisfying N^K, than it would be if it did not satisfy N^K. Unfortunately for Pointy, we have already seen that this is not true. If Pointy did not satisfy N^K, Pointy would be (identical to) Dull, and Dull is not a worse knife than Pointy. This means that N^K cannot be a norm governing Pointy, because Pointy is not better in virtue of satisfying N^K than it would be otherwise.

Having N^K play this constituting role at the threshold in the Pointy/Dull Saga threatens its ability to be a norm, because it threatens it's accordance with Norm Governance. If Pointy ceased satisfying N^K, it wouldn't thereby become a worse knife, it would (by hypothesis) cease being a knife at all. So, if norms are the standards that individuals would be better in virtue of satisfying and worse in virtue of failing to satisfy, N^K can't be a norm at all.[45] Norm Governance, which plays such an important role in our constitutivist sketch, is at odds with norms serving the role of minimal criteria of kind-membership. Norm Governance entails that it is possible to have individual kind-members that are negatively assessable according to every norm governing that kind. It is only by recognizing the possibility of such individuals that individuals in positions like Pointy's can be evaluated according to norms like N^K. Norm Governance requires that Dull and Pointy are both knives. So, the Threshold Commitment is at odds with Norm Governance.

This is an unacceptable consequence for constitutivists who take norms to evaluate individuals in virtue of their kind-membership. If Pointy is a knife and N^K is a norm of knives, then Norm Governance holds that Pointy is better in virtue of satisfying it and *would be worse* if Pointy did not. The Threshold Commitment is a way of giving a necessary condition on kind-membership in terms of norm satisfaction. But we have seen that this conflicts with a more fundamental commitment of the constitutivist: Norm Governance.

2.8 MOVING PAST THE THRESHOLD COMMITMENT

What we need is an account of the nature of evaluative kinds that links individuals, kinds, and constitutive norms that does not entail that those norms are necessarily satisfied. We started this chapter by looking at the appealing work that Functional Goodness allowed with artifacts and what made them evaluatively good *qua their kinds*. If we can understand evaluative kinds as having something like characteristic functions, then we could help ourselves to the Metanormative Commitment that characterizes constitutivism about norms.

But, of course, we need an account of what it is for any individual to be a member of

[45] Remember, this antecedent was an essential part of our constitutivist sketch in the first section.

a kind with a function, and this is where concerns about violability and bad actions were raised. It is a pervasive fear that constitutivsm cannot supply an account of norms that both explains what it is to be a member of a kind and also provides evaluative standards that individuals can fail according to. I showed that a natural (and common) move is open to constitutivists to avoid the Violability Worry. The Threshold Commitment was seen to allow individual kinds to be negatively assessable according to norms without threatening the violability of any particular norm. But allowing for functions to be separated from satisfying individual norms in this way has puzzlingly not moved critics. We saw that this becomes less mysterious when we see that the Threshold Commitment doesn't give a satisfying account of how individuals and kinds are related by norms. This unpalatable consequence will affect all constitutive accounts that appeal to the Threshold Commitment. If we want to ground norms in the nature of evaluative kinds, we need an account of kind-hood (functional or otherwise) that does not tie satisfying norms to being an instance of the kind governed by the norms. Any account that does so will conflict with Norm Governance, which is as an essential condition on norms as Violability.

I do think we can save the Metanormative Commitment, and doing so requires rejecting the Threshold Commitment. Such a move would require an account of kind-membership that did not make kind-membership a matter of meeting a minimal set of norms of that kind. Such an account should allow the Metanormative Commitment to generate norms without making fully defective kind-members impossible, thus avoiding the conflict with Norm Governance.

We can see that so far, I've shown how constitutivism is characterized by the Metanormative Commitment, but that any workable constitutivist account will require an account of conditions of kind-membership that allows for fully defective kind-members. I've argued that similar worries about violability and bad kind-members are easily met on current constitutivist accounts, and put it forward as a puzzle why this concern is so gripping. I argued that the view of kinds that is given by constitutivists is able to accommodate bad individual members of kinds and violable norms via the Threshold Commitment. The puzzle is answered by seeing that there is a problem with the Threshold Commitment; though it is able to accommodate bad individuals and violable norms, its denial of cases of extreme defect

causes a serious conflict with Norm Governance.

2.9 BYPASSING DISPOSITIONS

Norm Governance is essential for the evaluative aspect of norms, it's non-negotiable, and so the Threshold Commitment has to go. Some constitutivists, however, might be tempted to dispose of the Threshold Commitment as a condition on kind-membership and instead appeal to a condition of minimal dispositions. This variant condition would require being minimally disposed to perform the characteristic function, with the promise of avoiding conflict with Norm Governance. I call this commitment **Dispositional Kindhood**.

Dispositional Kindhood: It is a minimal necessary condition on an individual, x, being a member of an evaluative kind, K, with characteristic function, Z, that x is disposed to perform Z.

The hope is of the dispositional constitutivist is that we can distinguish functional kind-members from non-kind-members by appeal to dispositions. Dispositional Kindhood replaces the minimal satisfaction of norms of a kind with being disposed to perform the characteristic function of the kind. On such an account, being disposed to Z should not involve satisfying a norm of Ks, but it would be a condition on being a K-member. This is supposed to block the problem of having norms at the threshold serve as both minimal kind-membership conditions and evaluative standards.

Unfortunately, this is no more viable than the outright endorsement of the Threshold Commitment. In fact, given certain reasonable assumptions, it entails the Threshold Commitment. It replaces the minimal satisfaction of norms of a kind with being disposed to perform the characteristic function of the kind, but it turns out that on viable accounts of dispositions, being disposed to Z involves being minimally good at Z-ing. As norms are just those standards that satisfying makes a thing better as the sort of thing it is, being minimally good at Z-ing requires minimally satisfying norms. It thus conflicts with Norm Governance for the same reason the account that appeals to the Threshold Commitment for

kind-membership does: it entails the impossibility of fully defective kind-members.

To see why Dispositional Kindhood rules out the impossibility of fully defective kind-members, consider what the requirement to be disposed to Z amounts to. It doesn't entail the impossibility of full defect (the desired outcome) only if it is possible for an x to be disposed to Z and still negatively assessable according to every norm derivable from Z. The easy way to see that these two conditions are not co-possible is to see that the former requires being able to Z, and the latter rules out being able to Z. Being negatively assessable according to every norm of kind K is incompatible with being able to perform the characteristic function of K. The norms of Ks are those standards that satisfying would make the individual better at performing the function of K. An individual that failed according to all such standards would be constitutionally incapable of performing the characteristic function of Ks. But individuals that are *unable* to perform a function do not have a disposition to perform that function. Being constitutionally able to perform a function is a condition on being disposed to do so. So, Dull is unable to perform the function of cutting, and can therefore not count as disposed to cut. Fully defective kind-members must be kind-members, on pain of conflict with Norm Governance; but Dispositional Kindhood rules them out. So, Dispositional Kindhood, like the Threshold Commitment before it, conflicts with Norm Governance.

Fans of Dispositional Kindhood might object to my claim that being able to perform a function is a condition on being disposed to do so. Being disposed to Z in circumstances, C, is sometimes thought to require Z-ing in some instances of C, but masking is often taken to be able to account for how an individual that is otherwise disposed to do so to might fail to Z in C.[46] Those on the side of Dispositional Kindhood might hope, then, that masking can account for the fact that some K-members (e.g. Dull) are unable to perform their characteristic functions despite being disposed to do so. However, Dull's situation is not one that falls under commonly accepted masking conditions. Dull is not prevented from manifesting a disposition to cut because he is, say, wrapped in cloth or dipped in wax (a

[46]Briefly, a *mask* is a property an object can have that accounts for it not manifesting a disposition it has in a circumstance in which it otherwise would. So, a vase, even when well-packed, is fragile. But when carefully packed the vase will not break when struck. So, the property of being carefully packed is a mask for the disposition to break when struck. The masking property of being well-packed thus can explain why the vase did not manifest the disposition to break when struck, despite maintaining its disposition to do so. The term is introduced by Johnston (1992).

traditional way of making knives safe for transport). Dull is not guarded by security or in a world with nothing capable of being cut. In order to make sense of properties that count as masks for Dull we would have to countenance *intrinsic* masking.[47] Only internal, sustained, intrinsic masking would seem to provide us a way to account for how some object was constitutionally incapable of cutting, but still disposed to do so.

There are two problems with this suggestion. First, intrinsic masking is unpopular because masking is generally accepted to be an external phenomenon. So, while *being carefully packed* can be a masking property for fragility in vases, *being made of rubber* cannot. Being made of rubber is incompatible with fragility, and what accounts for this is, on many accounts, that it is an internal property.[48] Second, (and more problematically) even if masking properties could be internal properties, we haven't seen how fully defective kind-members could count as being disposed to perform their function, even admitting the possibility of internal masks. Full defect implies that there is no property that would further the function of that object. So, there is no property that would, absent the intrinsic mask, account for the manifestation of the disposition. Masking, intrinsic or extrinsic, still requires *other* intrinsic properties that account for the disposition that are themselves masked. However, there is no property that Dull has such that, in virtue of having it, he could possibly count as able to cut. This is just what it is to be a fully defective kind-member. Even were internal masks possible, there must be some property that accounts for the disposition, such that, if the internal mask were not there, the disposition would be manifested. But there is no such property for fully defective kind members.

Masking, whether intrinsic or extrinsic, will not help the fan of Dispositional Kindhood. Being disposed to perform function Z requires that there are some properties in virtue of which the disposition to Z would manifest itself in Z-ing. Masking can explain why, despite having those properties, an individual might not manifest the disposition, but masking

[47] A mask, dispositionally speaking, is a circumstance an object with a property can be in, such that though the object has the disposition to ϕ in circumstance C, the mask prevents ϕing in C. So, while a vase might be disposed to shatter when dropped, being carefully packed will mask this disposition. An intrinsic masking would thus be a circumstance that is an intrinsic property of the object in question, such that, though it was disposed to ϕ in C, there is masking condition that is intrinsic to the object that prevented ϕing.

[48] Intrinsic masking is not without its supporters. For arguments in support of intrinsic masking, see Ashwell (2010), Everett (2009), and Clarke (2010).

cannot account for how an object that has no properties that would further Z-ing could be disposed to Z. So, Dispositional Kindhood, like the Threshold Commitment, rules out the possibility of fully defective kinds. Both entail that individuals that are fully defective with respect to the norms N^K are not members of K. I've already argued that the Threshold Commitment should be abandoned because of this entailment's incompatibility with Norm Governance. Dispositional Kindhood should be abandoned for the same reason. Dull and its friends refute Dispositional Kindhood.

2.10 CONCLUSIONS

We started off this investigation with a popular yet controversial project for understanding norms: constitutivism. I put a diagnostic problem forward: people think constitutivism is plagued with problems (the Problem of Bad Action and the Violability Worry) that turn out to be easily resolved. Why, then, is this easily-solved problem so persistently *worrisome*? The answer was that there is an easy resolution for the stated worry (the Threshold Commitment), but it manifests with its own problems (so the worry never seems to be satisfactorily answered). It rules out fully defective kind-members and so conflicts with an essential aspect of norms—Norm Governance. This conflict explains why people view constitutivism with such suspicion, despite the ease with which the actual objections are answered. So, we vindicated the concern people have about constitutivism *and* explained the persistence of the worries: the way constitutivists account for kind-membership in evaluative kinds can't support a viable account of norms. I then tried one more possible answer that might be on peoples' minds, Dispositional Kindhood, and showed that it would have the same bad consequence of conflicting with Norm Governance.

I seem to have left constitutivism in a bad place. Constitutivists need a way to account for how individuals are members of evaluative kinds, because it is the nature of a kind that grounds the norms of that kind, but it is individual kind-members that are evaluable according to those norms. Two of what seem to be the most promising ways of giving necessary conditions on kind-membership no longer seem viable. To vindicate the constitutivist

project, the constitutivist has a project ahead of her: she needs to account for how norms are grounded in the nature of kinds (and thus validate the Metanormative Commitment) but avoid conflict with Norm Governance. Primarily, though, the constitutivist owes us an account of how individuals are members of kinds.

I do think we can save the Metanormative Commitment, and doing so requires rejecting the Threshold Commitment and thus dispositional accounts like Dispositional Kindhood. Such a move will require an account of kind-membership that does not make kind-membership a matter of meeting a minimal set of norms of that kind. With an account of the nature of evaluative kinds that avoids these threats, we can ground norms in those natures without making fully defective kind-members impossible, thus avoiding the conflict with Norm Governance. In the next chapter, I develop such a constitutive account of evaluative kinds as having characteristically functional natures without running afoul of Norm Governance

3.0 FUNCTIONS FOR CONSTITUTIVISTS

3.1 INTRODUCTION

In the first chapter, we began considering the prospects for an appealing account of the nature of norms—metanormative constitutivism. We were left in the difficult position of having to reject two of the more promising ways of accounting for the relationship between norms and individual kind-members, characterized by the Threshold Commitment and Dispositional Kindhood. The Threshold Commitment held that it is a necessary condition on being a member of an evaluative kind, K, that an individual minimally satisfy the norms of Ks. Dispositional Kindhood made it a requirement that an individual be disposed to perform the activity or function of Ks. Unfortunately, this strategy can't work. Appealing to norm satisfaction of the disposition to perform a function as a condition of kind-membership led to a serious problem for the possibility of evaluative norms. These accounts of kind-membership I showed to be incompatible with Norm Governance, which holds that norms that govern an object are those standards that the object would be better insofar as it satisfied and worse insofar as it didn't. As this is an essential feature of norms, these accounts can't be the place the constitutivist should look for the nature of kind-membership.

The constitutivist project aims to ground norms, i.e. evaluative standards, in the nature of the object to be evaluated. In the first chapter, we focused on artifacts and their status as evaluative kinds, which seemed tied to their having functions. In this chapter, I develop an account of evaluative kinds. I begin by considering accounts of function for biological kinds, where considerable progress has been made regarding how best to understand evaluative kinds as functional kinds. I'll endorse a (partial) reduction of biological evaluative kinds to functional kinds and then show how that functional nature can ground biological norms. I

then show how the account of biological function can be analogically extended to account for artifactual function as well, leading to the same grounding relation with norms governing artifacts. This will put us in a position to have a general account of the source of norms governing evaluative kinds. Absent a compelling reason to accept a disjunctive account of the grounding of norms, the constitutivist should want such a general account.[1]

If successful, we will have an account of evaluative kinds, in general, as functional kinds, which will then be able to ground norms. Our task is constrained by the known issues addressed in the first chapter. We need an account that doesn't prevent individuals that are defective according to a function from having that function, which would run afoul of the Metanormative Commitment. Also, it must be possible for the functions we care about to be had even by individuals that are fully unable to perform the function.[2] With such an account of function we will avoid the Threshold Commitment and thus not run afoul of Norm Governance. Locating such a function and explaining how it accounts for the nature of evaluative kinds will be the first substantial project of this chapter.

The project here is to answer the challenge to ground norms in the nature of the individuals evaluable by those norms. This chapter has roughly the form of a proof by cases: I will show that there is an account of function recognized in the philosophy of biology literature that, for biological kinds, can explain kind membership, functional nature, and norms governing those kinds. I will then argue that the account in the philosophy of biology can be generalized to a form that applies to the case of artifactual evaluative kinds considered in the first chapter and that this generalization should extend to all evaluative kinds.

[1] For arguments against the desirability and possibility of this kind of unity see Preston's "Why is a Wing like a Spoon?" These question are currently the subject of lengthy debate in the philosophy of technology literature, which space precludes addressing here. For a nice introduction to that debate, see Elder "A Different Kind of Natural Kind" and Vermaas and Houkes "Ascribing Functions to Technical Artifacts: A Challenge to Etiological Accounts of Functions."

[2] A discussion of why this must be so can be found in the penultimate section of Chapter One.

3.2 COMING TO KNOW AND LOVE PROPER FUNCTIONS

We're in the market for an account of function that can play a specific role in the constitutivist account. It needs to explain how individuals can have the function without entailing either that those individuals with a specific function are able to perform it or that those individuals meet any of the standards (i.e. the norms) that follow from having it. If such an account of function is unavailable (a risk we must acknowledge after the rough treatment such accounts received in the first chapter), so much the worse for what seems to be a promising form of general constitutivism about norms. Our the goal here is to see whether such an account is available: to see whether a plausible-seeming constitutivist account—grounding norms in functional nature—has the resources it needs to get off the ground.

In the philosophy of biology, *proper function* is the kind of function that something is taken to have non-accidentally, as a consequence of what that thing is.[3] A proper function is a one an object has because of its nature,[4] not because of its accidental or relational properties or circumstance. Linguistically, this difference between accidental and proper functions is marked by what are known as *strong function statements*, which pick out proper functions (e.g. "A has the function Z.") and *weak function statements*, which pick out accidental functions (e.g. "A performs function Z.").[5] When we claim that "The function of the heart is to pump blood," this is a claim that pumping blood is the proper function *of the heart*, where if I say, "That rock performs the function of a door stop," this weak function statement is a claim only that the rock can perform that function, not that the function is *the rock's own* or the functions of rocks, as such. This can't just be an appeal to linguistic usage (though it is that);[6] if there is going to be some account of function that is meant to distinguish this

[3]Proper functions are thus sometimes contrasted with what are known as *systems functions*, those that give a account of function of some x relative to the causal contribution x plays to the broader capacity of a system in which x can be understood as being a part. Cummins (1973) is the canonical reference for such a function, but see Allen et al (1998) for an overview of this distinction.

[4]Millikan stresses her coinage of 'proper' is meant to mirror its etymological ancestor, the Latin *proprius*, meaning 'one's own,' see her "Biofunctions: Two Paradigms," p116

[5]I've gotten this helpful terminology from Boorse. Weak function statements also include: "A serves the function Z," and "A functions as a Z-er." See Boorse "A Rebuttal of Functions." p 70 and fn10.

[6]Obviously, we can sensibly say "That rock has the function of a door stop," so it can't just be the felicity of assertions that is doing the work here. Rocks, as such, do not *as a kind* have the function of door stops, even if many of them are well-suited to serve such a function. It might be that some individual rock *does* have the proper function of a door-stop, but there would have to be some other kind to which the

part of the nature of evaluative kinds *as such* that can account for the norms that govern those individuals, it must be something like proper function.

Proper functions, those functions that an individual has as the result of its nature, are precisely the sort of function the constitutivist needs. In the philosophy of biology literature, etiological accounts of proper function—those that appeal to the history of individuals and kinds to account for function—are taken to be most able to account for the sorts of requirements the constitutivist has.[7] Consider the following quote from Millikan as she motivates an etiological account of proper function as accounting for *what it is to be a heart*:

> That a heart is a heart certainly has something to do with pumping blood. But what kind of connection with pumping blood must a heart have? Some hearts are diseased and some are malformed in such a way that they are unable to pump blood. Other devices, such as water pumps, are perfectly capable of pumping blood, yet these are not hearts... It is not then the actual constitution, powers, or dispositions of a thing that make it a member of a certain biological category. My claim will be that it is the "proper function" of a thing that puts it in a biological category, and this has to do not with its powers, but with its history.[8]

That some individual is a heart seems to have something to do with pumping blood, but why the function of *pumping blood* is part of the nature of hearts *as such* or how any individual object comes to have that nature is unclear. Millikan in the text above is, like us, in the market for an account of what makes an individual the thing it is, when this has something to do with *doing something*. We want to explain how individuals are members of evaluative kinds and, in a way, so does Millikan. Above, she notes that being any good at pumping blood isn't a condition on being a heart, which is a familiar point to those following along from chapter one. Moreover, some non-hearts might be pretty good at pumping blood if given a go at it, but aren't thereby hearts.[9] So being any good at pumping can't be what

rock belonged, in virtue of which it had that function, because rocks, as such, do not. It is not proper to something *qua rock* for it to have the function of a door stop. It is proper function that is tied to kinds and so can account for this distinction.

[7]Though I'll be unable to adequately address alternate accounts of function in this chapter, for a nice overview of the differences in etiological and propensity accounts of function, see Mitchell's "Dispositions or Etiologies?" For presentations and criticisms of the propensity view of functions, according to which the function of an individual or trait depends on how it would fare under selection in some specific environment, see Tinbergen (1963), Bechtel (1989), Horan (1989), Walsh (1996), and Walsh and Ariew (1996) and especially Bigelow and Pargetter (1987).

[8]Millikan "Proper Function" in *Function, Selection, and Design* p 85.

[9]Additionally, some things have been specifically *made* to be pretty good at pumping blood, but even

makes a particular heart *a heart*.[10] Millikan concludes that it can't be actual constitution or abilities (powers) or dispositions that make something a member of the biological kind *heart*. Instead, she claims we must appeal to the history of an individual and objects like it to account for their function.

If it isn't the powers, abilities, or constitution that can give something its functional nature, Millikan concludes it must be its history. According to etiological accounts of proper function, for some activity or effect, Z, to be the function of some individual, x,

i) x must be created in a process selected to produce individuals with a form F,

ii) F itself must have been selected over alternatives *because* it was selected to have effect Z.[11]

Millikan introduces the concept of 'reproductively established families' to pick out groupings of individuals that non-accidentally share a form due to being reproduced or copied either from each other (think: sexual reproduction), or by a common origin (think: protein expression from genetic code).[12] Reproductively established families are thus those groupings that satisfy the first criterion for having a proper function: they have been created by a copying or reproduction: reproductively established families share forms. Members of reproductively established families have a proper function when the production or replication

those things aren't *biological* hearts. For the sake of simplicity here, I've omitted this section from the Millikan quote. She is thinking about (roughly) biological organs, and it complicates the picture that there is a way to use the sortal 'heart' to pick out two kinds that have nested extensions: one that is the biological kind and another that encompasses both biological and artificial objects. It will turn out that both biological and artificial hearts count as having the proper function of pumping blood, but for the moment, I'm glossing over the fact that Millikan is interested here in specifically biological hearts.

[10] I am unable to locate anyone in print arguing that it is being good at pumping that could give something the function of a heart, but this position has been defended by several people in conversation. This position is one that it is having the use function of pumping blood that makes something a heart. Also, note that this sort of position, at least, the negative position, is fairly popular in neo-Aristotelean circles.

[11] In what follows, I primarily appeal in the text to Millikan and Neander, but the general etiological story is found in many other writers in the function literature. The differences among competing etiological accounts, for my purposes, are mainly differences in detail. For further specifications of what constraints etiological theories require, see, for example, Griffiths (1993) and Godfrey-Smith (1994).

[12] For Millikan's initial introduction of reproductively established families, see her (1984) Chs 2 and 3. She distinguishes two ways of being a member of a reproductively established family: being a member of a first-order reproductively established family and being a member of a higher-order reproductively established family. Briefly, first-order reproductively established families are reproduced via direct copying and second-order reproductively established families are reproductions in virtue of being created by a first-order reproductively established family. So, for example, reproductively established families of organisms are first-order, because sexual reproduction is a means of direct copying, while reproductively established families of organs are higher-order, because they rely on first-order families to carry out the reproduction processes.

process selected (together with constraints) that form F *because* it served or as expected to serve the function Z. Reproductively established families are groups of individuals that share a form non-accidentally, because they have a shared account of creation. So, when the form of the reproductively established family was selected for having an effect, that effect is the function of all members of that family. Not all reproductively established families have proper functions, because not all ways of being produced or replicated via copying involve constraints required for proper function. Members of a reproductively established family share a form as the result of copying and have a proper function when the form was selected in order to serve (or further) that function.

Proper function, thus, results from a specific sort of history: one that can account for why the kind of object in question has the form it does because that kind was selected to be as it is to serve that function. Having a proper function, then, is a matter of having been created in a way that relates the form with an end. So, for biological kinds, we have the beginnings of a sketch of how the constitutivist can explain how evaluative kinds are functional. What it is to have a proper function is to be a member of a reproductively established family whose individuals are created with forms in a way that accounts for those members having that form for the sake of an effect, which is the proper function. This is the etiological story of proper function. Neander writes:

> On an etiological account, the function of a token trait depends on the selective advantage of past traits of the type in ancestral individuals. Its function is to do what traits in the relevant lineage were selected for. So, before we can determine the function of a token trait, and therefore before we can classify it according to its function, we need to be able to locate it in a lineage. In general, to determine that a token trait, x, has the function, Z, we need to determine that x belongs to a lineage of traits that was selected for Z-ing.'[13]

Individual biological kind-members are individuals that result from selection pressure for a shared end or activity. The function of an individual cannot be separated from the function of the kind the individual is a member of, and that kind must have a history that picks out the function as being selected for. Neander's account, above, is specific to biological kind-members, but this kind of story need not be.[14] For the moment, I'll set aside

[13]Neander "Types of Traits." 403

[14]This sort of etiological account can be originally found in Wright's (1976), see esp. pp 81-2.

questions about generality.[15] Proper function, understood etiologically, does not require that the biological individual in question actually perform the proper function that it is selected to perform, and so it seems posed to avoid the violability worry.[16] This etiological story about proper functions will allow us to rule in defective and malformed hearts because of their membership in a reproductively established family of hearts.[17] So at least in the biological arena, something can have a proper function without being any good at performing it. This account, thus, groups kind-members according to what they *ought* to do, not according to what they *actually* do or are *disposed* to do.

This account of proper functions is relatively uncontroversial.[18] An individual has a proper function as the result of being a member of a reproductively established family whose form is selected to have a particular effect. The novel move is using such an account to show that evaluative kinds are evaluative *in virtue of* being kinds characterized by this account of proper function. Taking the non-functional description of kinds, the project aims to show that they have proper functions that reproductively established family members share, and that the kind these members compose is evaluable because it is functional in this way.

3.3 BIOLOGICAL EVALUATIVE KINDS AS FUNCTIONAL KINDS

Reproductively established families are not themselves biological kinds in our general conception, which encompasses organs, species, genes, traits, etc. Reproductively established families require shared history amongst family members, but biological kinds require no such shared historical account linking individual kind-members. In fact, it seems important to

[15]Millikan also does not restrict proper function to biological function "...consider the categories thermometer, can opener, heart, kidney... Anything falling in one of these categories is what it is, falls in the category it does, by reference to function. One way to focus on the problem that the definition of 'proper function' is designed to solve is to ask how individuals that fall under function categories are grouped into types." "In Defense of Proper Functions" p 21

[16]Remember, the violability worry was the worry critics of constitutivism raise that norms will be necessarily met by all things they apply to on the constitutivist account. Showing how current constitutivists avoid this worry was the first step in the argument of Chapter 2.

[17]It is unlikely that there is one unique reproductively established family of hearts, but a plurality. More on this to come.

[18]Though it is controversial that etiologically understood functions in this sense are what biologists care about. It's also controversial what the specific account of etiological functions must look like.

the goal of biological inquiry that this not be so. Determining shared biological kind is a way of recognizing commonality where one might not expect it, due precisely to developmental differences or lack of shared evolutionary history. We can *discover* that this bit of the newly discovered species is a liver, because it has the proper function of livers. Its presence is accounted for by selection pressures that selected its form to the perform the function of livers. This sort of explanation is unavailable to us if there is no possibility of shared biological kind-hood across reproductively established families.

Biological kinds, on this account, are groupings of reproductively established families according to shared proper function. Reproductively established families require shared history, but biological kinds, in the sense we are interested them as evaluative kinds don't. Though every heart is a member of *some* reproductively established family, there is not one reproductively established family of hearts. While all hearts might not be members of the same reproductively established family, they will all be members of *some* reproductively established family in virtue of which they have a proper function. The biological kind *heart* is then a functionally understood kind.

To see how this claim works, consider that two individuals are not necessarily members of the same reproductively established family in virtue of being members of the same biological kind (consider: bat wings and heron wings compose different reproductively established families, but members of both are members of the biological kind *wing*). The obvious move is then to see biological kinds as simply groupings of reproductively established families that share a common proper function. Because proper functions are multiply realizable,groupings of individuals that share a nature characterized by proper function will be group together distinct reproductively established families that share proper functions. So, biological kinds are essentially functional kinds that get their functions as the result of being or being composed of reproductively established families. Biological individuals thus would have functions in virtue of being created members of reproductively established families that have forms selected for that function, and share a kind with other individuals who have an analogous story.[19]

[19]One challenge to this line of thought is to object to grouping reproductively established families in this way to create "new" kinds. If individuals are members of functional kinds in virtue of their proper function we won't necessarily consolidate the number of kinds we have, we'll be populating our world with additional,

Consider the case of the biological kind of hearts. That something is a heart is accounted for by its having a proper function. The etiological story given above takes the story *through* reproductively established families, but what ends up making the heart *a heart* is having the proper function. Biologically, proper functions are effects for which traits were selected by natural selection, they are the final aspects given in the copy principle above—the activity or end for which the traits were selected.[20] Biological reproductively established families are the basic groupings of biological individuals. So, any number of reproductively established families can share the proper function of pumping blood. For any biological bit of the world, it shares its features non-accidentally with members of a reproductively established family and those members share any proper functions they have necessarily. But being a heart is not a matter of lineage, because not all hearts are related by lineage. Being a heart is a matter of having the proper function of pumping blood, and *that* is a matter of lineage.

It is only by recognizing biological evaluative kinds as groupings along shared function that we can, for example, make sense of the phenomenon of analogous trait kinds. Analogous traits are those that develop independently in different species for the same proper function but are not present in the last common ancestor of the species being considered. Because the selection process is not linked through copying non-accidentally (through lineage), analogous trait individuals do not share a common reproductively established family. So, while both bird wings and those of bats evolved from the vertebrate forelimb, biologists believe that the forelimb in the two species evolved into wings independently of their last common ancestor. Bat wings and bird wings thus form separate reproductively established families. But, because both reproductively established family had the form of the wing selected for its ability to provide flight, they share the proper function of flight. Wings are a biological kind, but they are not identical to any one reproductively established family. Wings, understood as a biological kind, thus has disjoint sub-classes of reproductively established families (including both bat wings and bird wings).

unnecessary, groupings. This seems wrong-headed to me. It is a mistake, I think, to hold that only the most fundamental elements of the world are real or worth considering. Hearts might be a disjunctive group of reproductively established families with the function to pump blood, but this does more than merely give us a useful way of conducting scientific research. It accounts for a real thing in the world: the assignment of functions. It marks a real division amongst things that exist.

[20]See also Millikan (1993) pp35-6, Neander "Functions as Selected Effects" p173 and "The Teleological Notion of Function" p459, and Godfrey-Smith (1994) p359.

It should be a surprise if two objects could not gain the same function in different processes, even biologically. It is open to us, for any member of the biological kind wing, to explain the norms that come from that kind in virtue of the function wings share, even if what accounts for an individual wing having the function of wings can differ by cases (due to different individuals being members of different reproductively established families). There's no reason why bat wings and bird wings could not both have developed the same proper function independently, forming different reproductively established families, but sharing the same proper function and thus constituting one biological kind: wing.[21]

3.3.1 Uniqueness Objection

There is a complication raised by taking biological evaluative kinds to be functional kind groupings in the way laid out above. We want individuals to be members of functional kinds in virtue of their proper functions, but biological kinds, as composed of sets of reproductively established families, can have multiple proper functions. This will lead to different groupings of functional kinds, depending on which function understanding us used for groupings. Berent Enç raises an objection along these lines with the example of the supposed function kind: eye. The mammal eye has a certain form: it has a lens that inverts light waves in the visible spectrum on the retina, which is a network of rods and cones that detect gradation in light and send signals conveying information about shapes, colors, and movement. Enç introduces the following problem with this proposed classification of the eye: Etiological accounts of function will generate two functions for the mammal eye: Z_1: to detect shapes, colors, and movement and Z_2: to detect gradation in light. This is a serial function (i.e. functions that are chained, where one function is selected for the accomplishment of a further function) because, Z_1 is accomplished because of Z_2.[22] But, if we think each functional kind must have unique proper function, and eyes are functional kinds, then whether mole rats and lizards, which have eyes of different structure, have *real* eyes, depends on which of Z_1 or Z_2

[21] When we later see that there are other ways of having proper functions in addition to being evolved biologically to have them, we will see (or, more precisely: we will be able to acknowledge what we already see) that there is a functional kind *wing* that includes both the biological kind and the artifactual kind that share the proper function.

[22] Enç gives this example in his "Indeterminacy of Function Attributions," pp294–295.

we choose. Enç writes:

> The mole rat's eyes are behind sealed eyelids and, by detecting levels of illumination, they cannot have the function Z_1. And if the *eye* is an organ the identity of which is determined by its function (and it is reasonable to suppose it is, for why else should two organs like the insect eye and the mammal eye, which have no structural features in common, be *eyes*), then the mole rat eye and the pineal eye [found in the lizard] are not real eyes.[23]

Enç takes this to be a problem: either eyes are functional kinds, understood as grouped under one *unique* function, or they are not. This is supposed to pose a dilemma for the account under consideration. If they are functional kinds, then there must be a unique function, but there seems to be no unique function grouping them all (and this seems to be imperialistic according to Enç) or there is no unique function, and we have not identified the biological kind with a function after all. Now, Enç doesn't consider the possibility that pineal and mole rat eyes are *defective* eyes. We can understand it as a fact about the etiology of the eyes of the mole rat—the family is such that they do not have Z_1 as a proper function, because their form (being behind sealed eyelids) was not selected to perform this function.

The problem seems to be that mole rat eyes and pineal eyes *are* eyes according to criteria including Z_2 but not according to those including Z_1. Because there seems to be no fact of the matter about which function eyes must have to count as eyes, we must either simply choose one of Z_1 or Z_2, which seems less a matter of how the world *itself* is and more a matter of how we are carving it up. Enç has found a more interesting case than he lets on, however, because pineal eyes are really quite strange. In fact, while it strikes me as infelicitous to say mole rat eyes are *not* real eyes, it seems equally infelicitous to say that pineal eyes *are* real eyes. Pineal eyes are formed during embryonic development from the cells that form the "real" (see how easy that is?) eyes of the lizard, but branch off and protrude from the top of the head. They detect ultraviolet light (thus satisfying Z_2) which triggers the darkening of the skin.

Here's one way to resolve the issue: Pineal eyes, I submit, are not true eyes, while mole rat eyes are, because there is a further unique serial function served by Z_2 in the cases of human eyes, insect eyes, and mole rat eyes, but not in cases of pineal eyes: Z_2 serves the purpose of Z_3: to provide phenomenal awareness of the surroundings. And phenomenal

[23] Enç "Indeterminancy of Function Attributions" p295

awareness of the surroundings via detection of light energy is *seeing*. Eyes are for seeing. Pineal eyes are not for seeing, they are for controlling pigmentation (what we might think of as Z_4). Mole rat eyes need to see much less, and insect eyes see with a very different structural apparatus. But eyes are functional kinds; they share the function of sight. Pineal eyes perform Z_1 not for the function Z_3, the function of eyes, but for Z_4. I don't have a sortal to pick out things with proper function Z_4, but perhaps there is one.

Now, this doesn't fully answer Enç's worry, because he seems to think that there must be a *unique* function for each functional kind. This might rule out serial functions even where the *final* function is shared across kinds, when some earlier functions are not. While human and insect eyes accomplish Z_3 via Z_2, mole rat eyes do not. They go from Z_1 to Z_3 without Z_2, because their sight is different than ours. So, if all three potential eye-types compose a functional kind, there will be different norms amongst kinds. What would be a defect in an insect eye (e.g. being unable to perform Z_2) might not be a defect in a mole rat eye.[24] Norms can thus be decided according to any proper function an individual has; knowing that something is an eye, thus, will be sufficient to determine *some* of the norms governing it, but to know whether it is functioning well or badly, one needs to know all the functions it has.

There are some objects that have multiple proper functions that do not have this serial structure, but instead have concurrent proper functions (i.e. multiple proper functions that are not related by accomplishment) but most are artifacts (e.g. things that Slice, Dice and Chop!).[25] In those cases, if reproductively established families *can* enable something to have developed with more than one proper function, they can be grouped into more than one functional kind. But this is not a problem. These cases will be ones which we recognize as legitimately cases of one object with two distinct functions (consider:. the hammer has

[24] Additionally, there is another sense in which all four eyes share a functional kind, though a functional kind we don't have a common sortal term for. They all have forms that have been selected to detect variation in light energy. So, they are all members of a kind *light sensors*. This seems true, and that some things that are light sensors are also eyes and some are not seems obviously true. Some things that are light sensors are artifacts designed for photographers. There's one in my phone that automatically dims the screen in low-light conditions and brightens it in sunlight. Neither of these things are eyes, but they share a function with eyes.

[25] It's possible that no individuals with concurrent proper functions are biological, given the way evolutionary constraints are thought to work.

the function both of pounding nails into walls and removing them from the wall and is not thereby mysterious).

This is all as it should be if biological evaluative kinds are to be evaluative in virtue of being functional. Reproductively established families are basic groupings of individuals that, when they have proper functions, have them in common with all other members of the reproductively established family, and biological kinds are functional groupings of reproductively established families according to their proper functions.

3.4 NORMS OF BIOLOGICAL KINDS

Biological evaluative kinds are, therefore, functional kinds, and proper functions are had by individual kind-members in virtue of their history and reproductively established family. Biological individuals can be grouped in various ways, according to their structural similarities or their reproductively established families, but insofar as the kinds we are discussing are biological evaluative kinds, they will be functionally understood. In showing that biological evaluative kinds are functional kinds, I have given the account of proper function that can make sense of this reduction and considered a few objections to this reduction. But this is all to the aim of showing that norms as such are grounded in the nature of their objects. The functional nature of biological evaluative kinds is the ground, but we still need to show that we can explain norms without running into problems like those canvassed in the first chapter.

I've introduced an account of etiologically understood proper functions for biological evaluative kinds, and shown roughly how they are supposed to work. The next challenge is to show how norms governing the biological evaluative kinds can be grounded in the functional nature of those kinds.

We've appropriated an account of proper function from biology that can account for how biological individuals have their proper functions (by being members of reproductively established families that have forms selected for an end) and how we can understand biological evaluative kinds in general as groupings of individuals by these proper functions. So, we

can see that biological individuals have proper functions as a consequence of their history. Fortunately, they turn out to help rather than hinder our understanding of the relationship between individuals and kinds. I've also addressed how this account allows biological kinds to include as kind-members individuals from distinct reproductively established families that share proper functions. So, we can account for how proper functions are gained by biological individuals—etiologically by membership in reproductively established families—and why biological kinds are essentially functional—they are groupings of reproductively established families with more or less specific accounts of proper function.

In the first chapter I argued that norms are standards governing objects that those objects would be better in virtue of satisfying and worse in virtue of failing to satisfy (characterized by what I called Norm Governance), and claimed there that properties that furthered fulfilling the function of the kind were virtues in individuals of the kind (what I called Good For). If proper functions can account for the functional nature of biological evaluative kinds, they are thus set up to account for virtues of kind-membership.

Etiological accounts of proper functions provide an account of what Millikan calls Normal conditions and Normal properties.[26] Normal conditions are those external conditions that the proper function was selected to be in and is thus generally required for performance of the successful performance proper functions. In cases of biological kinds, where proper functions are accounted for by evolutionary constraints, Normal conditions are those conditions that led to the selection and maintenance of the form for the function. So, being attached to the artery system, being in a body, having blood running through the arteries, being in normal temperature and pressure environment are Normal conditions for a heart. So, the heart won't pump if it's not getting electrical impulses, and pumping isn't going to accomplish circulation if the heart isn't in the body, and because not merely pumping, but also circulation, was selected for, the heart belongs in the body. Not because the heart is a better heart in the body, but because the heart's Normal condition is in the body. Normal properties are those properties required of the individual itself for performance of the proper functions. So, Normal properties for a heart will include being strong and structurally sound, having a

[26]I follow Millikan in capitalizing 'Normal' to separate it from 'normal' in the sense of 'average' or 'expected.' In many cases, Normal properties are held by almost no tokens of the kind, and so are not in the common statistical usage 'normal,'

certain capacity for pumping relative to the capacity of the circulatory system, etc.

Failures to have Normal properties amount to defect, they are vices, but failures to be in Normal conditions are not defects in an individual. After all, a heart outside of its Normal conditions might malfunction, but not because of defect, but because of circumstance.[27] In the course of serial functions, what I call the *teleological function* is the function under which all other proper functions in the series can be described. So, the teleological function of the heart is to power the circulatory system and the teleological function of the eye is to enable sight. The heart *pumps* in order to *move blood* in order to *circulate it* in order to *aid in distributing nutrients and collect waste* and etc. All of that is a description of how the heart mechanically powers the circulatory system. The eye *detects gradation in light waves* in order to *detect shapes, colors, and movement* in order to *provide sight.* Though it is (perhaps) possible that an object has more than one teleological function (see: Slicer, Dicer, Chopper!) this is only a challenge to the functional account of evaluative kinds if that makes it impossible to explain the norms that govern that object by appeal to its plural functions, and I can think of no reason that would be a worry.[28]

Normal properties seem to be exactly what we're looking for to ground norms. Normal properties are just those properties that norms of a kind would require on our account. Having them makes you better, qua your kind, and failing to have them makes you worse. Failing to have them all together does not threaten your kind-membership. There's no need to be at all disposed to satisfy the Normal properties of your kind to be a member of it.[29] The norms of biological kinds, then, can be grounded in the nature of the individuals governed by those norms. Biological kind-members have their functions in virtue of their history (the

[27]For discussion of the role of privileged function in distinguishing malfunction from failure to perform a function see Elder (1995).

[28]Again, it seems unproblematic to evaluate hammers according to two functions: it removes nails from the wall and it hammers nails into the wall. That some objects are complex functionally doesn't seem to threaten our ability to functionally assess them. This will be a topic addressed at length in the next chapter.

[29]But seeing how being disposed will come along with being a member of a reproductively established family is important. On this point Millikan writes "Having the *right sorts* of current properties and dispositions is in point of *fact*, in *our* world, an infallible index of having proper functions. If you like, it is criterial, as criterial, say, as the red of the litmus paper is of acidity. But it is not turning litmus paper red that *constitutes* acidity, nor is it having the right sort of current properties and dispositions that *constitutes* a thing's having a purpose. To the degree that each of these contemporary descriptions in terms of current properties or dispositions is successful, each describes only a *mark* or purposiveness, not the underlying structure." (1984) p19

biological case accounts for this as being members of reproductively established families), and are members of the biological kind they are in virtue of that function. But they're individually evaluable because their function can be furthered or frustrated by satisfaction or failure of norms that require Normal properties. The etiological story that accounts for the proper functions of kinds provides an account that explains what all objects that have a function share—they share having been created with a form selected for some activity and end(s).

Individuals should have those properties that further the performance of the activity and the accomplishment of the end: Normal properties. So, hearts should pump at a certain rate, because that is how the activity of pumping will best run the circulatory system. The teleological function is something that, in Normal conditions, an individual will be better or worse at achieving in virtue of having the Normal properties of the kind. So, it is according to the Normal properties that defect and not merely malfunction is assessed. It is Normal properties that an object is better in virtue of having and worse in virtue of lacking.

Constitutivists can explain the norms that govern biological kinds in terms of these Normal conditions and Normal properties. The norms of an individual with function Z are those standards that require Normal properties for Z. Normal conditions are those conditions that allow individuals of an evaluative kind to be better able to achieve the teleological function of the kind, but they do not license evaluation of individual kind-members. (A heart does not become a worse heart when the blood is drained from the body, but its Normal conditions cease to obtain, and thereby, the teleological function is not fulfilled). So, a heart ought to be strong (Normal Property), because this is how it can best aid the circulatory system. But it ought to be hooked up to the neural network (Normal Condition) because otherwise it cannot receive the electric impulses to pump, no matter how strong it might be.

3.5 GENERALIZATION TO ARTIFACTS

The task for the constitutivist extends beyond biological norms. The constitutivist project is to ground norms in general in the nature of the object governed by the norms. We've established that in the case of biological evaluative kinds, the individual biological kind-members have their functions as evolved objects, and that these functions serve as the determinant for more diverse groupings according to evaluative kinds. We account for proper functions of biological kind-members by recognizing their individuals as members of reproductively established families whose form is selected for a purpose by evolutionary constraints. We can account for those standards that require properties that would further the performance of those teleological functions to be norms of those objects (standards that require Normal properties). In biological cases, this selection is accounted for by evolutionary selection of adapted traits, but etiological accounts of function have structural elements which can be shared by other sorts of functionally-understood kinds. I will argue that this structural parallel obtains, and allows for a suitably unified theory of biological and artifactual functions.

The general ambitions of the functional grounding here demand an account that extends beyond the biological kinds. The hope is to generalize the etiological story of biological evaluative kinds to a story about creation for a purpose *in general*. This generalized story should account for proper functions in all evaluative kinds. We've established the case for biological evaluative kinds, and I hope here to extend the account to artifactual kinds as well. The role that evolution plays in accounting for the selection of forms for a purpose is taken to be accounted for by design in artifactual cases.

Biologically, the forms of reproductively established families are selected by evolutionary pressure from alternatives to have certain forms because those forms are suited (in Normal conditions) to perform activities for selected ends. Artifacts, in contrast, are designed: selected from alternative possible designs to have certain forms to perform or be suited to perform activities for ends. Tokens of both biological and artifactual types thus result from creative acts. Artifacts and biological kinds do share a kind of generality that can be captured by the shared application of copy principles: both are created kinds, though only one (artifacts) are created intentionally by a separate kind of entity (a designer) for

its own ends. Though only artifacts have *designers*, both artifactual evaluative kinds and biological ones have the form they do because that form has been selected for its suitability to a purpose.[30] Both artifacts and biological individuals are as they are because they do something. This is the benefit of the etiological account, design plays the analogous role in artifactual cases that evolutionary selection plays in biology.

This account of artifactual function is appealing over a rival account of how artifacts receive function that holds that artifacts have only "use functions," that there are no proper functions for artifacts, no privileging between the purpose the object was created to serve and the purpose an owner or user puts it to. On the position motivated in this chapter, in contrast, conditions of creation must obtain for an object to be subject to new evaluations as the kind of thing it is. It rules out a certain kind of power: normative omnipotence. We are not normative gods, we cannot make norms govern objects merely in virtue of our imposition of those norms.[31] Something isn't defective in virtue of not being sharp simply because I'd like to use it to cut. I might very much want to chop this sapling down with a rigid piece of aluminum siding, but the siding is not defective when I am unable to achieve this task with it. This is because the functions an object has are in general independent of my desires with respect to or appropriative uses of it. But we *are* normatively powerful in another sense: we are creative agents and we can imbue our creations with standards in virtue of this creative act.

Creation is a powerful type of act—it brings into being things that were previously not in existence. Creation is not merely ontologically powerful (i.e. by bringing into existence new individuals), but is normatively powerful as well—it can bring into existence a new kind of

[30]It doesn't even have to be suitable to have been selected, although this is a difficulty. There isn't a large literature on it, but there is an interesting question about how artifacts that have impossible functions could be created. These functions Beth Preston calls "Phantom Functions" and her example is to provide luck or ward off demons. So, amulets are things that have the function of warding off evil spirits (say). They are all bad at performing this function, because there are no evil spirits. But, amulets are still those things that have that function. Anyway, this is a particular issue for artifacts, and I think doesn't extend to other functional kind-types. But it is worth considering, because it seems like, on this view, any form selected to perform a function by a designer would make an individual functional kind-member, no matter how ill-suited to its function. Phantom functions make this issue more explicit—it shows how kind-members can be given functions that *couldn't* be performed at all.

[31]For an example of the position that select cases of this are, in fact, possible, see Ruth Chang's (2013). I don't have space here to argue against this position, but I think it is highly implausible and Chang's defense is limited to the creation of reasons, not normative standards in general.

item that is governed by norms where there were no such objects before. The number and kinds of things governed by norms increases as the result of such acts. We can understand the power exercised in creative acts that are normatively powerful according to the same copy principle we used to understand biological kinds as being composed of reproductively established families. Creation, though powerful, need not involve an intentional creator of a different type than the object created. (Even if you are not yet on board with the implications of this line of thought, it should be uncontroversial that biological reproduction is an act of creation, and need not involve a creator in any other role than the reproductive ancestors.) Things that have their own principles of self-governance that involve reproduction or self-propagation can be credited with creation of tokens of their own type in the sense relevant here. They involve the copying of their own form when this can be explained because the form in question is selected to perform certain activities for certain ends or aims, and those ends or aims are the proper function of the copied individual.

So, creation is a type of act that can bring tokens of functional kinds into existence. Appealing to the etiological account of creation according to a copy-principle to account for being a kind-member can explain how both biological and artifactual individuals can be functional kind-members. But I've only gestured to how this general story can account for both biological and artifactual norms, in virtue of the functional nature of these evaluative kinds. Both biological kinds and artifactual kinds are pretty unproblematically functional. If you think any kinds have functions, you'll think biological and artifactual kinds do. I've done some work to show that those functions are held even by individuals who are incapable of performing the function and that norms can be explained in terms of those functions, but this doesn't accomplish showing that norms governing evaluative kinds, as such, can be explained in this way.

3.6 EXIT CONDITIONS

There's a final concern I'll address here about using etiologically understood proper functions to account for the nature of evaluative kinds. We've seen that proper functions can be

used to provide an account of biological and artifactual evaluative kinds that accounts for them as functional kinds. But, it doesn't seem to provide an explanation of how functional individuals could *lose* their function. There is a worry here about entry and exit conditions: being descended from a set of individuals produced in a certain way or being created with a form selected for a purpose is an *entry condition* on kind-hood, but this seems not to leave any room for *exit conditions*. Having a history doesn't seem to be the sort of thing that can be lost. If having a function is a matter of having a certain history, then it seems that there can be no exit conditions on functional kind-hood, which is a serious problem. Lineage is something that cannot be lost, but function certainly seems to be. Let's call this the Problem of Exit Conditions.

The central role of proper functions in kind-hood cannot account for exit conditions if we think that history is the only thing that gives a collection of matter its function. If that's all it is to have a function, then it seems that the pile of wood that was once a chair (before the circus elephant stomped on it) would still be a chair (after the stomping). This is one motivation for the Threshold Commitment considered and dismissed in the last chapter: it is implausible that there can be no conditions under which something could cease to have a function it once had, and the Threshold Commitment accounted for these conditions by the position that there is some limit condition of norm satisfaction that served as such a minimal condition. We saw that we must reject the Threshold Commitment as a minimal condition on kind-hood, but we should recognize that those who endorsed the Threshold Commitment had one thing right: it's implausible that there could be no way to cease being a kind-member.[32]

In the case of biological and artifactual kinds, individual kind-members have functions as the result of being selected to have a form for the sake of a function. In the case of biological kinds, this is accounted for by membership in a reproductively established family whose form was selected by evolutionary constraints to further the fitness of the kind. Relevantly, members of reproductively established families come into existence through an act of creating with a form. Similarly in the case of artifacts—they have functions when they have been created to have a form selected by a designer to perform a function. In both cases the

[32]For this argument, see especially sections 4 and 5 of the first chapter.

functions and the form come together in an act of creation. Because of this component of creation, there will be some conditions under which they should go out of existence—conditions under which they would be *destroyed*. How an individual could lose a proper function, then, is only as complicated as how one could cease existing.

Artifactual and biological kinds are physically created, and thus destruction would be physical. Because an object (like the chair) can be destroyed without destroying the matter it is composed of, persistence conditions of objects is a thorny problem in metaphysics. Putting this aside for the moment, there is an intuitive difference between *damaging* an object and *destroying* it. If you destroy it, there is one fewer thing in existence, where if you damage it, there are the same number of things, but one of them is worse than before. The Threshold Commitment puts damaging and destroying on a continuum, where destroying something is just sufficiently damaging it. But this, of course, can't be right; there is more to destroying than damaging severely enough. This is because damaging (as I'm understanding it here) is a matter of interfering with the Normal properties of the object. If I cut through this stack of newspaper with this knife, I will damage the knife by dulling it. It will be duller and therefore less able to perform its proper function of *cutting*. It will manifest the Normal property of sharpness to a lesser degree. But to destroy the knife is not to merely dull it, I must physically alter it in a way that is more than merely reducing the Normal properties it manifests. If I break the handle off the blade, *perhaps* I have then destroyed the knife. Notice, this turns on whether it is a Normal property of knives that they are whole. If it is a virtue in a knife (rather than a condition on being a knife) that it is unified, then by breaking it into parts, I have damaged it, not destroyed it. It seems to me that it is not a Normal property of knives that they are whole, it seems like a constitutive fact about knives that they are whole. But we don't need to settle the issue; there will be some cases where we might not be sure if the issue is one of destruction or damage. With physical objects, both will involve physical acts, so ambiguity might be unavoidable.

This kind of ambiguity is what motivates the Threshold Commitment linking of damage and destruction: being sturdy is a virtue in houses, Korsgaard argues, and sufficient vice in the realm of sturdy-ness leads to physical destruction: so, you must be minimally structurally

sound to be a house.[33] But this doesn't prove what she thinks. Consider: a totally unstable house can be held up by external scaffolding, such that it would fall and be destroyed if the scaffolding were removed, but it is still a house, despite itself having *no structural integrity*. Structural integrity is the internal assurance of unity. But it is *unity* that is destroyed in destruction, and a unified house is not a *better* house, because, as we saw already, there is nothing for a unified house to be better than, there are no dis-unified houses. Korsgaard is right that unity is essential to kind-hood, she is just wrong about the relationship unity has to norms.

What counts as physical destruction, rather than damage or mutilation, is perhaps a matter of vagueness because what counts as being unified can be a matter of vagueness. Between any two clear cases in which one leaves a damaged kind-member and the other destroys it, we're likely to find an instance where there is a less clear fact-of-the-matter case. There must be something that counts as destruction, and it cannot be accounted for by counting up normative violations. What it must be, at least in the case of physically created kinds, is physical destruction. I don't have much to add to help clarify what amount of unity is needed for destruction, but this is a problem for anyone who thinks that the objects we interact with on a daily basis were created, exist, and can be destroyed. Persistence conditions for objects are a shared problem.

3.7 CONCLUSION

So much space to make constitutivism plausible, and I've avoided all mention (more or less) of the main bread and butter of constitutivist accounts: practical and theoretical reason. Constitutivists get the attention they do because they propose to tackle norms governing actions and mental kinds in ways that resolve classic problems in action theory and theory of rationality. They propose to make those norms essentially binding on their objects, they argue that it is essential to the very nature of agency or rationality as such that rational agents are bound by the norms they are. But what I've developed so far does not provide

[33] Korsgaard (2009), Ch 2.

an account that gives such a story.

In the next chapter, I will start to give such a story. Actions and mental states in fact *are* evaluative kinds in the sense laid out so far. They are created kinds that have their norms as the result of proper functions that come from the copy principles they are established according to. For now, though, notice that we've given the constitutivist what she needs. We've provided an account of function that allows violability, normativity, and (for now) generality. I've argued that the notion of proper function is suited to provide the constraints needed to allow for the and normativity and shown that the account generalizes to artifacts as well as the biological sphere where proper function has its first home. It is less obvious that the account generalizes even further to account for the spheres of practical and theoretical evaluative kinds, but the task of the next chapter will be to motivate this further expansion of the account and begin to defend it from some initial objections.

4.0 CONSTITUTIVISM AND UNITY

4.1 RECAP AND SETUP

In the first chapter, we began considering the prospects for an appealing account of norms—Metanormative Constitutivism. The account proved appealing because it grounds norms in the nature of evaluative kinds, those kinds whose members can be better or worse, *qua* member of their kind. But in Chapter One we were left in the difficult position of having to reject two of the more promising ways of accounting for the nature of evaluative kinds.[1] Grounding norms in the nature of evaluative kinds remained appealing, however. In the most straightforward cases (artifacts and biological kinds), it seemed clear that the norms in question were straightforward evaluations of the individuals *qua* members of their kinds that, when satisfied, made those individuals better as a kind-members.

At the end of the first chapter, we saw that an otherwise attractive way of understanding evaluative kinds wasn't workable, and were left without an account of the nature of evaluative kinds that would allow us to derive norms from that nature. Chapter Two began the task of developing an account that could do this work. There, I developed an account of evaluative kinds that could ground norms using an account of proper function adapted from the

[1]They were characterized by the Threshold Commitment and Dispositional Kindhood. The Threshold Commitment held that it is a necessary condition on being a member of an evaluative kind, K, that an individual minimally satisfy the norms of Ks. Dispositional Kindhood made it a requirement that an individual be disposed to perform the activity or function of Ks. Unfortunately, this strategy couldn't be made to work. Appealing to norm satisfaction of the disposition to perform a function as a condition of kind-membership led to a serious problem for the possibility of evaluative norms. These accounts of kind-membership were incompatible with Norm Governance, which holds that norms that govern an object are those standards that the object would be better insofar as it satisfied and worse insofar as it wasn't. As this is an essential feature of norms, the Threshold Commitment and related positions can't be the place the constitutivist should look for the nature of kind-membership.

philosophy of biology.[2] A proper function is a function an object has because of its nature,[3] rather than because of its circumstances or its accidental or relational properties. Proper functions turned out to be precisely the sort of function the constitutivist needs to explain norms. In Chapter Two, I argued that only an etiological account of proper function—one that appeals to the history of individuals and kinds—is able to account for the constraints raised by explaining how functions could be responsible for evaluative norms.

Etiological proper functions require a specific sort of history: one that accounts for why the kind of object in question has the form it does by appealing to its selection history. The individual must be a member of a reproductively established family whose form was selected because it served a particular function. Having a proper function, then, is a matter of having been created in a way that relates the form to an end; it's thus teleological. So we have the beginnings of a sketch of how the constitutivist might explain how evaluative kinds are functional kinds. To have a proper function is to be a member of a group whose individuals are created with forms in a way that accounts for those members having that form because the form was selected because it produced an effect, which is the proper function. Proper function, thus understood, does not require any particular individual to actually perform the proper function that its form was selected to perform, and so it is poised to avoid the violability worry.[4] This etiological account of proper functions also allowed us to, e.g. rule in defective and malformed hearts because of their membership in a reproductively established family of hearts. This account, thus, groups kind-members according to their function, what they *ought* to do, not according to what they *actually* do or even what they are *disposed* to do. All it is to be a member of a functional kind, to have a proper function, is to have an etiology that accounts for the proper function of the kind.

The move in Chapter Two began by taking account of an individual or group in question

[2]Proper functions are sometimes contrasted with what are known as systems functions, those that give a account of function of some x relative to the causal contribution x plays in the broader capacity of a system in which x can be understood as being a part. Cummins (1973) is the canonical reference for such an account, but see Allen et al (1998) for an overview of this distinction.

[3]Millikan stresses her coinage of 'proper' is meant to mirror its etymological ancestor, the Latin *proprius*, meaning 'one's own,' see her "Biofunctions: Two Paradigms," p 116.

[4]Remember, the violability worry is the worry critics of constitutivism raise that constitutive norms will be necessarily met by all things they apply to on the constitutivist account. Showing that current constitutivists avoid this worry was the first step in the argument of Chapter One.

that does not presuppose functions, including the story of its creation (either evolutionarily- or artificially-understood), and from that non-functional account derived a function that explained the norms that govern individuals of that kind. In cases where there is an etiological proper function, that function could be used to group disparate collections of created individuals (i.e. different reproductively established families) into a single evaluative kind according to a shared function. These groupings according to proper functions are evaluative because of their shared function. Chapter Two thus concluded by showing that artifacts and biological kinds can be understood as *evaluative* kinds because they are *functional* kinds. The norms of these kinds could be grounded in their functions, because they are standards that individuals that satisfy them are better at performing their characteristic functions. The account of etiological proper functions allowed us to explain, for the relatively simple cases of artifactual and biological norms, how those norms were grounded in the functional nature of the evaluative kinds they govern.

The task of this chapter is to first motivate and then make good on the promise of unity of norms that motivates the constitutivist project developed in Chapter One and Two. This requires first motivating the appeal of unity by answering some initial objections to the possibility of the unity of norms governing such disparate kinds as artifacts, organs, reasoning, and action. I then respond to some *a priori* worries that threaten to show that the etiological functional account cannot, in principle, yield a unified account of norms. In the three primary sections that follow, I'll address worries that threaten these goals. In 4.2, I consider objections to unity in general and motivate a unified account on independent grounds. In 4.2.1, I argue for the desirability of a general account of norms grounded in the nature of the objects they govern. In 4.2.2, I consider whether differences in epistemic access to the objects of different norms threatens the possibility of unity and argue it does not. In 4.2.3, I discuss the relationship between linguistic norm use and the objects of norms, which if it is too close threatens to undermine unity because of differences in the application of norms in different areas. I argue that this is not a threat to unity. 4.2.4 and 4.2.5 conclude a response to these objections and locate some plausible candidates for the objects of theoretical norms. In section 4.3, I consider some general *a priori* objections to the grounding of norms in etiological function. In 4.3.1, I consider a form of well-known

counter-example, Davidson's Swampman. I discuss objections that are often taken to show that some individuals or kinds can be evaluative without having etiological functions and some that are taken to show that some functional individuals are not thereby evaluative. In 4.3.2, I address worries ("Martian Worries") that grouping evaluative kinds according to history might rule out some groups because of a lack of shared history. I argue that this rests on a misunderstanding of the role etiology plays in this account. I conclude the chapter in 4.4 by discussing the way that unity of norms provided by the metanormative account on offer can avoid general *a priori* worries, but admit that empirical worries are still a possible threat to the account's plausibility.

4.2 DISUNITY WORRY

One of the motivations of the Metanormative project that is the subject of this dissertation is the thought that all norms, i.e. all standards that license evaluations of essential goodness or badness, are related to their objects in a particular way that is the source of those evaluations. This claim has, at its heart, a commitment to the unity of evaluation.[5]

When I say that 'x is a good K-member,' there is one way of relating x and Ks that can tell us whether there is some norm that licenses this evaluation. The way that x became a K-member might vary depending on the K in question, but there is a commonality in all essential goodness and badness claims. At least, this will hold in all such claims that are true.

There are several ways that the case of theoretical norms differs from the cases of artifactual and biological norms. In this section, I'll begin by addressing the most pressing worry—that unlike in the case of artifacts and biological kinds, theoretical norms do not lend themselves to the attributive evaluative assessment of their objects in the form 'good

[5]I think there are (at least) two ways of having a unified constitutivist project: the rationalist way and my way. The rationalist finds unity in the source of all norms: rationality. The metanormative constitutivist, in contrast, thinks that all norms have one unified explanation: their relationship to their object. The objects are different, but the relationship is the same. The rationalist view has what I take to be the devastating consequence that in order to have any norms at all (and thus anything that is better or worse), you need rational beings. It seems unlikely that you can explain, for example, what makes something a better or worse ficus by appeal to what is constitutive of rationality.

K' for some obvious K. While the norms of biological kinds lend themselves to sorting things into good and bad organs and traits and species members and the norms of artifacts lend themselves to sorting things into good and bad knives and corkscrews and circuit breakers, theoretical norms do not lend themselves to this attributive usage. We end up with normative conclusions like "so-and-so is rational" or "this is good evidence for that" which are not clearly translatable into some good-making feature held by some evaluative kind. That is, they do not wear their evaluative-kind relatedness on their sleeves. This might threaten to undermine the essential unity between norms and evaluative kinds and their functions that is being exploited by the metanormative constitutivist.

4.2.1 Why Unity?

According to the etiological account of proper function, an individual x has proper function Z if x is a member of group G and Gs are created with form F *because* F was selected because it enabled Gs to Z. Take two cases: First: eyes. An individual organ has the function of seeing if it is a member of a group of organs that were created with a form that was selected because it enabled sight. 'Eye,' I claimed, is just a sortal we use for things that have the function of seeing. When you find a group of things that have a form selected to perform the function of seeing, you've found a group of eyes. Well done, biologist you! Similarly: knives. An individual object has the function of cutting if it is a member of a group that is created with a form that was because it enabled cutting. Those things created with a form selected to cut are knives. When you find a group of things that have a form selected to perform the function of cutting, you've found a group of knives. Well done, archeologist (or shopper) you!

In these cases, we start with some object and ask "what is this thing?" and come up with an answer. "It's an eye!" ("All the better to see you with...") "It's a knife!" ("Buy it! We need to chop some onions!")[6] The case of theoretical norms and their objects is different,

[6]It is, on my account, possible that the relevant information needed to *determine with certainty* what any individual object is may be removed from us. Archaeologists, on this account, can never likely be certain what any artifact *really* is, and thus their confidence in the application of norms to their finds will be less than if they had purchased the artifact at a store. Because I see no reason to expect the nature of objects to be luminous to us, this does not strike me as a problem for this view.

because we are first confronted with *norms*. For example, we are told "you should apportion your beliefs in accordance with the evidence" or "do not reason fallaciously." In this case, the direction of our investigation seems reversed, we are confronted with normative commands rather than the objects of normative assessment. In the cases of artifacts and biological kinds, we move from the morphology of something to its etiology to its functional kind and *then, at last,* we can derive the norms from that function. But the cases of theoretical (and practical) norms seem different.

In the case of these norms, we don't have the same access to the objects we want to investigate. They don't lend themselves to empirical observation and manipulation in the way artifacts and biological kind-members do. It seems there is a problem supplying a unified account of norms. In the case of artifacts and biological kinds, for each kind, K, K's nature grounds facts about good-making features for K-members. And standards that require those good making features are norms of Ks. But if this worry can't be answered, it seems like there is no K that theoretical norms pick out ways of being good *qua* that we can appeal to. Without access to such a kind (or kinds), grounding theoretical norms, the unity that is one of the motivating features of the metanormative account being developed here is undermined.

4.2.2 Epistemic Access Disunity

Because we don't seem to have direct access to the objects of theoretical and practical norms in the way we do to (most of) the objects of artifactual and biological norms, there's a worry that we can't have a unified account that relies on such access. Thankfully, though it's indirect, I think we do have access to what the objects of theoretical norms are.

Though we do not determine the objects of theoretical norms empirically, I think we can determine what those objects are by finding the most plausible norms in the area and what they actually evaluate as good or bad. I suggest we can determine what the objects of these norms *must be* by seeing how we go about *applying* them. Though this means that we have a stronger grip on the norms than the thing assessable by the norms, this doesn't undermine our ability to assess our intuitive norms according to the nature of the objects they govern

and revise them. We have two tools to identify the object from the norm.

First, we can rule out certain items by recognizing them as things that aren't plausibly the sort of thing that could satisfy the norm. That is, we can notice that the corollary of *ought implies can* is *cannot implies not ought*.[7] Second, we can see what it would be intuitively appropriate to assess as better or worse if the norms at issue were violated. Because norms are those standards that would, when violated, make an individual governed by them worse (and better when satisfied),[8] we can consider what would thereby be worse when the norm is violated. Intuitively, that thing is the object of the norm.

So, though our access to the evaluative objects of norms in the case of theoretical norms differs from the cases of artifactual and biological norms, it doesn't seem to threaten our ability to locate the objects in question. We're simply in a different epistemic position in relation to these objects, because they are not immediately presented to us in the way artifacts and biological individuals are.

This move might seem too quick, however. Moving from the norms to the individuals that would be better or worse according to them only works if there is some basis for agreement on theoretical norms to begin with. There is hardly universal agreement on what the right theoretical norms are and the more commitments we take on with respect to these disagreements, the less generally appealing the final account will be. This could be a problem because the metanormative constitutivist account promises to help settle some of these disputes. Starting from theoretical norms and moving to their objects threatens to give this benefit up by requiring us to settle difficult disputes between participants in the literature over theoretical norms for seemingly little pay-off.

I think this objection would be a serious problem for the metanormative theorist if there were considerably more disagreement about theoretical norms. But as it turns out, the disagreements tend to be over things like what the force of theoretical norms is (e.g. over whether they truly normative in some robust sense) or how to formalize them (e.g. whether wide scope or narrow scope deontic operators are the best modal representations). There

[7]Of course, there are readings of 'ought implies can' on which things I argue in in Chapter One imply that ought does not imply can; but I mean to pick out a sense in which 'x ought to y' implies that x is the sort of thing that can y. This sense will work just fine here.

[8]For a reminder of why this gloss of norms is an appealing one, read the section introducing Norm Governance as a constraint on our understanding of norms in Chapter one.

is a fairly substantial amount of agreement among those who disagree about these aspects of theoretical norms. For example, people agree that belief formation processes ought to be truth-conducive and reliable, that we ought to apportion our beliefs to the evidence, etc. Of course, there are legitimate disagreements over which of these are primarily important or linked to knowledge or justification or what evidence is. None the less, there is broad consensus that these are important normative requirements in the theoretical domain.

4.2.3 False Linguistic Unity

While the last worry suggested that beginning with norms can't satisfactorily determine the object of norms, we might instead think that in the case of theoretical norms the object is just as clear as in biological or artifactual cases. The reliance on attributive good in the constitutivist account might lead us to think that there is some obvious K we can find in our linguistic practice surrounding theoretical norms that would be 'good K's when those theoretical norms are satisfied. We might think, for example, that theoretical norms are just the norms of theoretical rationality and that it's almost analytic that the object of the norms of theoretical rationality is theoretical rationality itself. Just like knife norms govern *knives*, the thought goes, that the norms of theoretical rationality govern *theoretical rationality*. If this is true, we don't need an investigation about what objects the norms govern; we just need to study theoretical rationality directly, in just the way we go about studying knives. It's not that moving backwards from the norms to their objects doesn't work, it's that it's no move at all.

Unfortunately, the linguistic similarity is superficial. Theoretical rationality is either a laudatory normative success grouping (like "healthy")[9] or it is an evaluative kind that can support norms (like "liver" or "human being"). It is a corollary of the main conclusion of Chapter One that no kind grouping could require minimal normative success and be an evaluative kind supporting its own norms. So, it can't be theoretical rationality in the

[9]I call these terms, those that pick out those individuals that satisfy some threshold condition of norms, 'questionable kinds,' I do not think they are always problematic or to be avoided (for example, 'healthy person' is a perfectly unproblematic questionable kind), but we should be careful to avoid taking them to be the object of our evaluative norms. This topic is beyond the scope of this dissertation, but I have a manuscript exploring this topic.

laudatory sense that is evaluated as better or worse when theoretical norms are satisfied or not. Could the object of theoretical norms be an evaluative kind of theoretical rationality?

Though we might try to be very careful, rejecting the threshold condition linking the two senses of 'theoretically rational,' having it both ways doesn't seem to help clarify the situation. I think that there is a simple way to see that we should privilege the laudatory sense over the evaluative kind usage. Even if we take the route of naming the thing evaluated 'theoretical rationality,' we still have to go about figuring out what *that* is. And it's not clear we have a great idea of that, or even that there's one unified thing that it picks out.[10] So, I will use 'theoretically rational' in the laudatory sense—as a way to pick out those individuals (or attribute to those individuals having the property of) satisfying theoretical norms particularly well.[11]

This leads us into a related "too easy" worry about locating the object of these norms. There seems to be an obvious grammatical subject of many of our theoretical normative evaluations: people.[12] The grammatical object of our admonishment when a theoretical norm is violated is often a person. We might, for example, judge that Jessie is irrational if she affirms the consequent, doesn't consider base rates in calculating probabilities, or rates conjuncts as less probable than their conjunction (in short, if she's your average undergrad in a psych study). So, we might take this to indicate that Jessie is being evaluated as bad when we say she is irrational. But, what kind of thing are we implicitly evaluating Jessie *as* when we say that she is bad in virtue of violating these norms.[13] Jessie doesn't seem to be a worse *person* because she affirms the consequent, and this is what would follow if it was *qua* person that we were evaluating Jessie. She might be (or be disposed to be) a worse *reasoner*, but that seems just to say she's bad at making inferences.[14] And you are bad at making

[10]Though I choose the laudatory use over the kind use, for or present purposes either choice is fine, so long as you only pick one and don't think that by picking the evaluative kind sense you've solved any of the problems.

[11]There will be a further distinction we'll look at later between taking the laudatory use to pick out the proper functioning of inferential capacities or as satisfaction of what we take the theoretical norms to be.

[12]We also can evaluate inferences themselves, but which is prior is not important here. If you don't think that people are the primary subject of theoretical normative evaluation, a variant of this objection might still appeal to you. If not, feel free to skip ahead.

[13]Remember, from Chapter One, strictly speaking, one shouldn't say that Jessie is being evaluated as bad *simpliciter*. In order to be evaluated as bad, one must be evaluated as bad *qua* member of some kind *for* for some use.

[14]It's slightly more difficult, here, because we want to block worries about lucky successes—Jessie is not a

inferences, at least in part, when you make bad inferences (or you refrain from making them when you should, but this isn't Jessie's sin here). In such a case, your capacities aren't ones that enable you to make good inferences consistently. It's then the mental capacities that are bad, and Jessie reveals herself to be a bad reasoner by making the bad inferences. So we can only evaluate Jessie as a bad reasoner with another object that is bad, in this case, the inferential capacity.

4.2.4 Theoretical Objects as Evaluative Kind-members

The prospect of unity turns on there being, for all evaluative kinds, a nature that grounds the normative evaluations of members of that kind. Of course, we can use the norms derivatively to assess persons (like Jessie), but it isn't Jessie that is bad in this characteristic way. So, if not theoretical rationality or persons, then what could be the object of evaluation of theoretical norms? The strategy was to see what could be assessed by a set of norms *and* be better or worse *as the thing it is* when it violates them. The thing that plays both those roles is the object of the norm, that thing whose function will determine the norms that govern it. We saw that there might be some disagreement about what norms require, or how they should be formulated, but we're going to see how far we can get without settling those questions. We can't settle the question entirely here, of course, but I think we can get part of the way to understanding what objects are implicitly governed by our intuitive theoretical norms.

When we say that in affirming the consequent Jessie is irrational, she might thereby be a worse reasoner, but this is because there is something *bad* about her exercise of her inferential or judgment capacities. It cannot be Jessie, directly, who is bad in virtue of this failing. Rather, it is her mental capacities (like those of belief-revision and other belief-formation capacities) that are thereby bad. Accordingly, the sets of mental states that those capacities take as inputs might also be the objects of theoretical norms, if they have functions as well. So, for example, we might say that Jessie is irrational when she affirms the consequent (i.e. uses bad belief forming techniques), believes both that p and something

great reasoner if all of her bad inferences somehow *cause* the world to accord with them.

that entails not p (i.e. has inconsistent beliefs), or thinks that she is far more liked than she is (i.e. has delusional beliefs).[15] These all seem to be ways of violating norms of theoretical rationality. In each case, a norm is or has been violated. She reasons contrary to "don't reason from 'q' and 'p then q' to 'p'." Or she doesn't take base rates into account when calculating probability. Or she rates conjuncts to be less probable than their conjunction. Or she doesn't apportion her beliefs (e.g. "Everyone adores my contributions to seminar") to her evidence (e.g. of people groaning and rolling their eyes when she speaks). The norms here have these objects: sets of mental states and mental capacities to perform inferences and make judgments. If we can see what accounts for these things having a proper function (if they do), we can figure out what makes them better or worse.

The thing (or things) evaluated by theoretical norms is not an organ or an artifact, but it could still have a proper function. It seems that there are roughly two types of norms of theoretical rationality, and there's a bit of a turf war about which are primary, which are important, which are really normative, which have the right form, etc. The disagreement about the form of rational requirements seems to track a disagreement about whether the correct theoretical norms evaluate inferential capacities (which would evaluate, perhaps, belief-forming processes) or belief sets (which would evaluate, perhaps, mental states). It seems, then, that these norms govern different objects. The capacity (or operation of the capacity) is bad when you violate a process requirement norm; the overall mental state is bad when you violate a state norm.[16] You can make better or worse inferences or you can be in a better or worse epistemic state; just as you can be bad at selecting knives (or making them, if you are a knife-maker) or bad at keeping the ones you have in good condition. Because inferential capacities and epistemic states are both theoretical matters, failures in any of these regards are grounds for being assessed as theoretically irrational. But none of this gives us reason to doubt the possibility of a unified account of norm-governed objects.

[15] I think it's more likely, actually, that having inconsistent beliefs is not itself a bad-making feature of a rational capacity, but perhaps is a sign of such a bad-making feature.

[16] Kolodny (2007), Broome (2007), Reisner (2009) all discuss the relationships between state and process requirements.

4.2.5 Open Issues

So, we have identified objects of evaluation that have etiological creation stories: inferential capacities. To confirm that the theoretical norms we take to govern these individuals are legitimate norms, we would need an account of the origin and history of our inferential capacity and belief sets to see what the proper functions of these objects are.[17] With an understanding of their proper function, we'd be able to determine those standards that select properties that would be virtues in individuals with that function. So, if we're able to understand these kinds as functional in the right way, we can verify that the theoretical norms we take to bind them can be explained by appeal to their proper function.

Giving the details of such an account is beyond the scope of my current project. I am just in the business, here, of showing that such an account is not ruled with some immediate objections. To do so I need to respond to several general challenges to this extension of the constitutivist project. Here we just need to make it plausible that the objects we evaluate with norms of theoretical rationality have a functional nature that can support those norms in the way the constitutivist claims.

4.3 COUNTER-EXAMPLES AND THE A PRIORI

4.3.1 Swampcreatures and Missing Etiology

> Suppose lighting strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, The Swampman... moves into my house and seems to write articles on radical interpretation. No one can tell the difference.
> But there *is* a difference. My replica can't recognize my friends; it can't recognize anything, since it never cognized anything in the first place... I don't see how my replica can be said to mean anything by the sounds it makes, nor to have any thoughts.

[17] It is, perhaps, worth adding a note on my use of 'individuals' and 'objects,' It has been a source of tension with some readers and interlocutors to use these terms to refer to things as diverse as capacities, inferences, persons, and organs. I take these terms to be metaphysically flexible here. If you find these terms problematically theory-laden, you can replace any instance of them with 'thing,' I simply find the term 'thing' too inelegant. I don't think substance is lost, however, in this substitution.

–Davidson[18, 19]

> There are no answers to a vast number of questions about Swampman that can be answered about every human, such as when he was born and where, who his parents were, which is his native language, and so forth. Swampman has no CV. More crucially, no predicates that make reference to normality or abnormality, even of the statistical kind, apply to Swampman, and none that make reference to the proper or improper functioning of his body parts or brain. He is neither tall nor short, neither strong nor weak, neither quick nor slow, either physically or mentally, for each of these predicates makes reference to one's biological group and he has none. None of his organs are functioning either properly or improperly except relative to an observer's interests. If his brain makes good cannibal soup, that is as much its function as thinking. Since he has in his head no organs that are adapted for concept development, belief fixation, contradiction elimination and so forth, nothing specifies whether he uses his brain correctly or incorrectly. This, I suggests, entails that he thinks neither truly nor falsely, indeed, that he thinks not at all.
> –Millikan[20]

Davidson's case raises the worry that Swampman's ontogeny, his intellectual development, was wrong. Specifically, it was lacking. Swampman couldn't *recognize* anything, because he hadn't *cognized* anything to begin with; his writings and thoughts couldn't mean anything, because they didn't have content on Davidson's account. Millikan, alternately, is concerned about his missing phylogenesis; Swampman has no supporting evolutionary history to make, for example, his glossy forehead protuberances *eyes*. Swampman's missing etiology undermines our ability to apply theoretical (or any) norms to him. Millikan and Davidson are committed to etiological accounts of mental states as well as organs. Most importantly, following our metanormative account, we endorse the conditional that if an object is norm-governed, then it is a member of an etiologically understood functional kind.

Though the larger projects of Millikan and Davidson commit them to these etiological accounts of mental states, qualia, and even consciousness, I want to take a moment here to show how my position diverges from theirs. Consciousness, I hold, does not have a proper function and thus it need not have history. Similarly, neither must having qualia or experiential content. These things I think *are* functions. Things can be selected for because

[18]Davidson (1987) pp443-4

[19]Millikan herself raises the example of randomly occurring duplicates and points out that they would "have no ideas, no beliefs, no intentions, no aspirations, no fears, and no hopes... This because the evolutionary history of the being would be wrong." in her 1984 *Language, Thought and Other Biological Categories*. I use the Davidson example because it is recognized as canonical, not because it was first.

[20]Millikan "On Swampkinds" p110

they provide consciousness, consciousness itself isn't a form that is selected, it is a function that an individual form can be selected *for*. So, there's no reason to think that the linking of etiology and function and norms that is under consideration here need threaten Swampman's ability to have consciousness, beliefs, or qualia.

Nonetheless, whatever the object of theoretical norms, Swampman doesn't have an etiology that can provide functional kinds to support them. Swampcreatures thus seem to provide a ready objection to the metanormative project of the constitutivist. The metanormative constitutivist seeks to explain which norms govern individuals by reference to what the functional natures of those individuals are. These functional natures, I've argued in Chapter Two, must appeal to the history of the individual in question to explain how it is a member of a functional kind. This account is undermined if there are norm-governed individuals that lack any evolutionary or developmental story. Swampman's glossy forehead protruberances seem, intuitively, to be eyes; he seems to make inferences in "his" post-swamp writings. But on the account here, it seems that either these seemings are mistaken or eyes and inferences, contra common sense, can't be better or worse.

4.3.1.1 Non-Answers to SwampObjections

Swamppeople are possible examples that make salient that basing norms in etiological functions rules out norms governing individuals that have no appropriate history. There's no way around this. Swampcreatures don't have histories and so don't have any functional parts. If Swampcreatures such as SwampDavidson are possible, it shows that there are possible objects created in wonky ways depriving them of histories that can support a function in the etiological sense. So they can't be governed by norms or be better or worse as the sort of things they are.

Some of the more common responses to Swampobjections aren't going to be helpful to us here. For example, those who are unconcerned with essences think Swampobjections "cast the net of counterfactuality" too widely in an attempt to tell us about the essence of things.[21] If you don't care about essences, and care only about how things are in *this* world, restriction to actually existing Davidsons should take care of the Swampcases.[22] Unfortunately, essences

[21] See Dennett's "Cow-sharks, Magnets, and Swampman" for this anti-essentialism.

[22] However, for a particularly gripping variant that makes it salient how *this* might be the world of Swampman, involving a case where you are asked to imagine yourself as Swampman (in a Brain in a Vat *how*

are exactly the sort of thing we're after here. Though I haven't framed the issue in terms of essences, asking after the nature of evaluative kinds, as I am doing, is asking after the essence of evaluative kinds. If there are logically possible evaluative kinds that do not have functions, as swampcases seem to threaten, this is a problem.

Alternately, Millikan and Neander have raised doubts about the nomological possibility of Swampcreatures. Millikan, in her "On Swampkinds" gives a nice discussion of how unlikely Swampman's creation is (not just statistically, with the lightning strike and all). She writes that "you can't in principle build large organic molecules such as hemoglobin... just by throwing the right amino acids together at the right angles with the right energies and having them stick. The trouble is that they *won't* stick!"[23] Neander also uses the improbability to bolster the argument against its use as a counter-example.

> The biologists are not entitled to reject Swampcow just because she is nonactual, and nor are we entitled to ignore Swampman because he is nonactual. More to the point, it is no accident that they are nonactual. The odds against Swampman or Swampcow are staggering. It is an extremely deep fact about the actual world and about all nomologically nearby possible worlds that highly complex organized systems are created by natural selection (or perhaps by beings who are the product of natural selection). Swampman and Swampcow, even if they did exist, would be freaks of nature in a fundamental sense. If we are trying to carve nature at its joints—or more modestly, if we are trying to develop the most useful theoretical notions that we can—we could discover that this deep fact provides useful constitutive conditions for biological and psychological kinds.[24]

I'm unswayed by both of these responses because I don't think the likelihood of Swampcreation has much to do with the force of the thought experiment. I'm not interested in carving existing (or likely existing) nature at the joints, I'm interested in carving nature at the joints. Our task here is not accounting for how existing things have their functions, or even developing the most useful theoretical notions for our world (this might reasonably be thought of as the biologists' task), it is to account for what must be true about something for it to be a member of an evaluative kind. Hypothetical possible things count, too. Our account can be threatened by merely possible things, so the more modest account isn't enough. For this reason, the staggeringly small odds of Swampcow are no consolation. It would be a saving

could you know scenario) see Ludwig (1996). And for a humorous but pointed take on what a world with Swampcreatures would be like see Louise Antony's "Equal Rights for Swamp-persons."

[23] Millikan 1996 p115

[24] Neander "Swampman Meets Swampcow" pp124-5.

grace if it turned out that hemoglobin required historical processes to be formed by amino acids "sticking" in just the right way, but we couldn't be guaranteed that that historical process was tied to a creation process in the right sense to guarantee a functional evaluative kind. We are bound to be stuck with possible objects that are not themselves function and yet are exact physical replicas of individuals that *are* functional kind-members. The question is whether this is a threat to our account.

We need an explanation of our intuitions and a set of judgments that would serve to evaluate, in some sense, these functionless individuals. Though the shockingly low odds of functionless duplicates doesn't prevent the objection right away, these odds can help explain our intuitions. Dretske, in "Absent Qualia," introduces an useful explanation of our intuition that Swampman has a heart is for pumping blood. He introduces what he calls the Paley Syndrome as "an irresistible tendency to use resemblance and placement as a basis for attributing purpose and design."[25] He points out that there doesn't seem to be anything irrational about manifesting Paley Syndrome, because it is a perfectly respectable form of analogical reasoning. However, in cases where we are not in normal (or, shall we say, Normal) circumstances (say, if we are confronted with Swampman), it may lead us to attribute purpose and design where there is none. Dretske claims that the syndrome is such a part of our inferential practice that "we use it to generate intuitions about cases even when we 'know' it doesn't apply."[26] So, though the staggeringly small odds that Swampcows exist (or are likely to exist in near-by worlds), doesn't give us reason to privilege a simpler account of functions, but it might help us provide an error theory for some of our intuitions about Swampcreature norms and functions.

4.3.1.2 Error Theory of Swampnorms

What we need, then, is an error theory of swampnorms. Dretske's Paley Syndrome seems to give us some account of why we might mistakenly think that there is a shared history or function between things with shared form. This sets us up for an account of our intuitions about Swampcreatures that might save the project. But a mistake about shared history or form does not give us the main thing our

[25] Dretske "Absent Qualia," p81
[26] Dretske "Absent Qualia," p82.

intuitions seem to demand: that we are permitted to apply norms to these Swampcreatures, even when we concede, explicitly, the lack of design history.

Neander introduces a Swampcow named Craisy, a swampreplica of the normal-historied cow Daisy. Neander claims that while Craisy cannot be a biological cow (because she lacks ancestral relationship to actual cows), she is, as Neander calls it, 'cow-like,' Neander writes

> Her actual dispositions can approximate the idea as well as any real cow's can: she still has something morphologically identical to some normal cow's heart that is still genuinely pumping something morphologically identical to some normal cow's blood. If history is needed to tell us what a bovine heart is supposed to do, it does not follow that history is needed to tell us whether Craisy's replica-heart does it.[27]

So, even if it's true that Craisy does not *have* a heart, Craisy has something that *performs the function* of a heart. To re-introduce terminology from Chapter Two: Craisy does not have a heart, a thing with the proper function of pumping, but she does have something that has a system function of pumping. Everyone agrees that Craisy has something that is pumping blood around her body; this is the function of hearts. What is at issue is whether the thing that is pumping liquid in Craisy *is* a heart. The question isn't what *is* going on, the question is what *should* be going on. We want to make cow-centric-assesments of Craisy. Is Craisy healthy? Is her blood pressure too high? Is the thing that's pumping malformed or diseased? These are questions that reference to mere replica-hearts can't answer.[28] It's not enough that there are system functions afoot. Chapter two convinced us that the only way to get what we want from functions is to have proper functions, those ones that don't commit us to a threshold.

Neander seems to think that Craisy doesn't have any relevance to biology or zoology. But Craisy does seem like a threat to our account of how we intuitively apply norms. If Craisy were to suddenly die because the liquid that was carrying oxygen to her tissues stopped being pumped (i.e. if the system function of pumping stopped being performed), we might be tempted to say something (her heart!?) malfunctioned. But if the etiological

[27] Neander "Swampman Meets Swampcow" p122

[28] Neander seems to think that because zoologists aren't threatened by Craisy that we ought not be threatened by SwampDavidson. I'm not sure I see how this argument goes, other than that because zoologists do not share the intuition that Craisy's organs have functions or that Craisy is a cow or any of the other problematic intuitions philosophers are supposed to have about Swampman that philosophers should give up our intuitions.

account of proper function is right, appeal to nearby actual cows with the same form isn't enough to get this claim of malfunction. We need *her* replica-heart to have the right history in order to account for it being the sort of thing that can malfunction in the first place. An event we can correctly characterize as a pumping-stoppage happened. But whether we can correctly identify it as a heart malfunctioning requires more than reference to a morphologically identical object that *would* be malfunctioning if *it* acted that way. We can certainly ask of Craisy's replica-heart how effectively it's pumping the replica-blood. (Alternately: how well it's performing the system function of pumping.) But if we want to apply norms, we need to appeal to the function of that replica-heart. We only get to apply norms if the nature of the thing in question supports those norms. The Craisy example is engineered to exclude the nature but impel us to still think the norm application is legitimate.

However, I think it is a mistake to think that we need to apply norms to get many of the evaluations that we want to make of Swampcreatures. Remember, norms govern evaluations of good *qua member of a kind*. They evaluate things *in virtue of the kind of thing they are*. This is the sort of evaluation that is denied to the replica-eyes of SwampDavidson and the replica-heart of Craisy. But it is not the only kind of judgment we can make. We can also make assessments that something is good *for* some purpose. And there are lots of purposes things are being put to in both SwampDavidson and Craisy. Lots of things are happening, and those things can be good or bad for lots of purposes.

SwampDavidson, despite having no organs and an amazingly functionless body, has lots of system functions being performed by parts of his body. It seems unproblematic that his replica-heart is only a make-shift heart and that his replica-memories aren't really memories; they are doing an great job performing the function of a heart and memories—they have these as system functions, if not proper functions. All of this is something we can make judgments about.

I think we have two kinds of intuitions about Swampcreatures. First, they are doing the same sorts of things as their non-Swamp counter-parts—looking around(both), writing articles(SwampDavidson), chewing cud(Craisy), etc. Second, they are assessable for doing those things better or worse. What we want, remember, was to make normative judgments about Swampman and Craisy, but the lack of function was stopping us.

I think SwampDavidson and Craisy are insanely lucky. They've lucked into things that will do the job that other organisms had to suffer for (literally, I guess). SwampDavidson has replica-eyes that are doing a pretty good job (as good as Davidson's) at what SwampDavidson is using them for: seeing. They're pretty good for seeing, even though that does not make them eyes. We can say "SwampDavidson has pretty bad eyes," when all I can mean was that his forehead protuberances are pretty bad at serving the proper function of eyes.[29]

It's counter-intuitive that a replica-heart isn't defective when it stops pumping abruptly. But, like Neander, we can still say that Swampman's replica-heart isn't performing the behavior that Swampman needs it to perform in order to live. It is bad for Swampman's continued survival. Given Dretske's Paley Syndrome, it's unsurprising that this very unlikely scenario elicits counter-intuitive results.[30] What rules out something *being* a heart (in this case: lacking a history) need not rule something out *serving the function* of a heart. And in this case, that will account for the counter-intuitive sense of this result.

No one is arguing that Swampcows and Swamppeople lack the capacity to pump blood. What is at issue is whether they are the sorts of things that ought to, as what they are. It is true that historical explanations do not accompany all possible capacities, but historical explanations distinguish mere capacities, or system functions, from proper functions. Because there are things that pump, we can determine when they do a better or worse job at pumping, but this can't make them better or worse hearts. Assuming that Swamppeople have interests (following Antony) it might be in their interests that their replica-hearts pump better.

As I mentioned earlier in introducing Swampcreatures, though Millikan and Davidson hold that Swamppersons can't have beliefs or interests, I see no reason why we can't think that Swamppersons can have these basic mental states. We are committed to the position that these kinds can't be evaluative or functional. If I were a Swampperson (following Ludwig (1996)'s suggestion to consider the possibility), I wouldn't care very much whether I had a heart or a replica-heart, so long as it performed the function of a heart. This just shows

[29]For this distinction, see Ch 1 of this dissertation and Ch 1 of Thomson's *Normativity*.

[30]Do we have grounds to rule out Swampman's thoughts as thoughts and Swampman himself as a person or should we, as Louise Antony calls for us to do, "stop stigmatizing spontaneous coalescence"? For Antony's eloquent appeal in defense of those who originate from the Swamp, see her "Equal Rights for Swamp-persons."

that in some cases, we care about things that are in our interests, not the actual functions of things. But that's consistent with how the world is. What we care about need not always track how the world is, independently of our cares. The two can come, conceptually, apart, and when they do, we might only care about our interests or the interests of things with interests.

The more Swampman is like Davidson (the example, by design, has him being *exactly* like Davidson), the more he triggers our Paley abductive inference patterns—sameness of form = sameness of design. Swampman is a case designed to keep internal states and external dispositions the same, but eliminate history. And our intuitions that such a being, who had all the same internal states and externally manifested dispositions, is supposed to pull us to see that we can't account for what something is by appealing to its history. But the fact that we generally infer shared essences by resemblance (which is accounted for by similarity among Reproductively Established Kinds) does not mean that sameness of resemblance actually perfectly reliably tracks sameness of essence. I don't mean to reject our intuitions on the grounds that Swampman is so unlikely.[31]

4.3.1.3 Swampjudgments [32]

Tolerate one more Swampcreation. Dretske has an example in his (1996) of a replicated Toyota Tercel that is almost exactly like *his* Tercel, except where Dretske's Tercel has a working fuel gauge, the Replica Tercel has something resembling a fuel gauge that is unresponsive to the amount of fuel in the tank. Dretske points out that the replica Tercel has no fuel gauge and so can't have a broken one, because it was not designed. There are things we can say about the replica Tercel, though. We might still wish that the replica-gauge were responsive to fuel levels. And so, we might be within our rights to make it responsive to fuel levels (say, by bringing it to a mechanic). We can say "Hey, you know what would make this Replica Tercel better at getting me around town? If this thing here were responsive to fuel levels."

[31] Of course, whether Swampman *can* have interests without having an etiology is partially what is at issue.

[32] The 'swampX' locution might here imply that 'swampjudgments' are seeming-judgments which lack proper etiology. I take Swampjudgments here to be judgments of swampcreations, rather than judgments from the swamp..

In a sense, Swampman is both the Replica Tercel *and* the owner of the Replica Tercel. In failing to pump blood (analogy: fail to respond to changes in gas levels) the replica-heart isn't defective, but Swampman might very much want it to continue to pump, and go out of his way to see that it doesn't stop, by, say, eating a healthy diet and exercising regularly (analogy: bringing it to a mechanic).[33]

So, Swampcreatures do not threaten our ability to make certain assessments about things of swamp origin. They are merely examples of things that cannot be better or worse *qua* what they are. SwampDavidson is *not* Davidson, and so we can't say the same things about him that we can about Davidson. But we can say some really similar things, and the things we can't say we can't say for good reason. Even though we can't make assessments about how Swampman ought to be, given what he is, we can assess how certain things going on with him ought to be, given what Swampman wants and needs. That is, given that one can have ends without having a history, we can assess what would make it better *for Swampman's ends*. Swampman is a very strange sort of being. Swampman is like a conscious being that lucked out in having a perfectly organized mess of functionless parts which perfectly perform desired system functions.

Consciousness, having qualia and experiential content, and other mental phenomena, I argued in 4.3.1, do not have proper functions, rather they are proper functions. The parts of our brains that provide *us* these things are functional. Swampman is someone who has lucked into consciousness; he didn't get it by design or by evolutionary trial and error. None of his parts have the proper function of providing consciousness, or giving beliefs, or qualia. But assuming he has them all the same, Swampman shows us that abilities and capacities can be, at least it is logically possible for them to be, randomly acquired.

We can't negatively assess Swampman's parts qua what they are, but we can assess them qua Swampman's interests.[34] If Swampman is a strange (highly improbable) sort of thing

[33]Of course, this assumes that Swampman can have psychological states, but we've seen that there's nothing on my view that rules out Swampman having wants and beliefs, just that they can't be better or worse *as such*.

[34]I'm assuming that Swampman can have interests. I think that, if Swampman is a Philosophical Zombie, the example loses much of its interest. However, this is a contentious claim, as many teleological accounts will want to hold that the intentionality necessary for Swampman to have interests is barred by his lack of history. I think that if he has replica thoughts, these are enough to get him interests, and if he lacks replica thoughts, I'm not inclined to have intuitions that he is the same as Davidson dispositionally speaking.

who has interests without having functional parts that provide them, Swampman is a strange hypothetical creature. He's not a human, but he is able to do and experience all that humans can. Swampman shows us that not only does something having a proper function not make it the sort of thing we want around (think: atomic bombs), but being the sort of thing we want around does not require having a proper function (think: Swampman's replica-heart).[35] Moreover, as soon as we recognize this about Swampman, it becomes pretty understandable how we might come to speak and think as if his parts do have functions, because they serve functions they do not have. Many of our intuitions about Swampman are not judgments of goodness or defect, but about what is good for Swampman. These generally come together, but Swampman shows us they need not.

4.3.2 Martians and Divergent Etiologies

While the worries above address not being able to account for intuitive norms applying to possible objects of evaluation that lack any etiology, we might also worry that the way that the unified account here is supposed to work will rule out some intuitively possible objects of theoretical evaluation because they don't have the proper etiology because of the work shared history seems to play in the account. By tying norms to shared history, we seem to risk ruling certain types of individuals out of evaluative kinds on the wrong sort of grounds. So, for example, if you think it's an open question whether or not there are rational bonobos or Martians, you might think it is unreasonable to settle the matter by whether or not they share a history with people. I like to think of this as the Martian Worry.

So, the objection is that because the view here ties being governed by norms to being a member of a reproductively established family and that this makes being governed by norms too tied to historical factors. Because Martians, if they exist, do not share an etiology with human beings, it would seem that they cannot be governed by the norms of theoretical rationality.

This objection turns on a misunderstanding of the role history and reproductively established

[35] Why care about proper function, then? Reread Chapter 3.

families play an important role in the account here, but they do not hold norms and functions hostage to sameness. A history and a reproductively established family are conditions on having a function which in turn grounds normative standards. But it is a fundamental aspect of the constitutive account that functions are multiply realizable. Different etiologies and different reproductively established families can share a function on this account, and if their mental capacities have an etiology (even one that differs dramatically from ours) and share the function that characterizes theoretical rationality (whatever it is) then they will be governed by the norms of theoretical rationality.

If Martians exist and if they have an etiology that accounts for a proper function that makes them reasoning beings (if we can call the kind of thing subject to theoretical norms reasoners), they are thereby *non-accidentally* governed by the theoretical norms that we are.[36] That we are reasoners is independent of the fact that they are, and vice versa, but *qua* reasoners, we are both governed by the same norms, and non-accidentally so.

Evaluative kinds, on my account, are the groupings of individuals (via their historically grouped reproductively established families) that can be better or worse qua members that kind under some description. It was the task of Chapter Two to show how we could understand evaluative kinds as grouped by proper functions. Etiology serves to account for how these proper functions *are had*, not to separate evaluative kinds. So, it is an open question who belongs in the evaluative kind reasoner. It's a matter of which reproductively established families there are with etiologies that can support the relevant function. Perhaps Martians belong, perhaps dolphins, perhaps angels.[37] All you need is the function, and that can be accounted for by many and varied etiologies.

The Martian Worry overlooks an important benefit of the constitutivist picture. In this account, you must look to the function that the form of reproductively established family of the individual was selected for. In cases where you find functions that are selected for etiologically, you have found a group of individuals with a function. The threat of the Martian Worry rests on the mistake of thinking that any one such group must be the unique

[36]It might be a matter of contingent historical events that both Martians and human beings are reasoners, but it is not an accident that they are both, as reasoners, both governed by the norms of theoretical rationality.

[37]In fact, the Indian government recently declared that dolphins are non-human persons, and extended certain protections to them. Though I don't think they've determined this based on an etiological function, if dolphins are reasoners, they certainly have such a selection story.

group of individuals with this function and thus must be the unique group governed by the resulting norms.

In the case of artifacts it is clear that this is not how function assignments work etiologically. Finding one group of things that has the etiological function of cutting does not speak at all to the possibility of finding other groups of things that do not share the same etiology but do share a proper function. We could, for instance, discover Martian knives, developed completely independently of our knives here on Earth. We can find different knives from different unrelated ancient peoples, with totally different artifactual etiologies. These knives could none the less all share norms because they all share the function of knives.

A similar response is available to us in the case of groups governed by theoretical norms, reasoners. Reasoners require an etiological function that is criterial of them, but they do not require a shared etiology. As should now be familiar, the same etiological function can be achieved through many and various etiologies. It is shared etiological *function*, not shared etiological *account*, that is needed for shared proper function. So, it is open to us to find reasoning Martians (come on, Curiosity Rover!). It is possible that Neanderthals were reasoners, despite the possibility of their developing reasoning capacities independently from Homo Sapiens. Any of Octopuses, Bonobos, Dolphins, or Bees could be reasoners. This account does not rule out the possibility of deities or angelic beings (which it might, if you think such things are necessarily reasoners if they exist). None of these things are ruled out on grounds that they do not share etiologies with human beings. Because it is shared proper function, not shared etiological account, that is required for shared norm-governance. My account does not hold theoretical norms hostage to facts about humans that are not shared by all reasoners.

Thus, the account here doesn't fall to either of the *a priori* worries addressed in this section. Though Swampcreatures' possibility tells us some strange things about how eyes and mental capacities can be gained, we are able to account for our intuitions about these cases and provide an evaluative language. Though Swampman is not a reasoner or a possessor of a liver or eyes, we can account for why it seems that he is, and through Swampman's interests, we can assess the performances of his parts as good or bad for his own interests, though they are not thereby better or worse qua any evaluative kind. Addressing the Martian Worry

allowed us to see that groups with distinct etiological accounts can still share functions and thus form a single evaluative kind. We are thus not forced, on this account, to rule out the possibility of non-human reasoners.

4.3.3 Venutians and Divergent Functions

There is a residual anxiety we might have about the possibility of reasoning Martians. Even if Martians with different etiologies might still, in virtue of their etiology, share proper functions and thus theoretical norms with us, there is a risk that some minded aliens might not. That is, if having mental states doesn't require any particular etiology, then there might be another kind of alien, let's postulate these are Venutians, who have the same mental states we possess, and who seem to have the same mental capacities, but who have wildly different etiologies that support wildly different functions and norms.

We seem, then, to be committed to saying of the Venutians that they have totally different norms for their identical mental capacities, even though they have the same mental states, capacities, and perform the same behaviors. Worse, because the only content that we have given to "theoretically rational" is success according the norms which govern mental capacities, it seems that these Venutians could be rational when their mental capacities performed in ways that, for us, would be highly irrational. For example, it might be that when they are making highly unreliable inferences, they might not thereby be irrational. This seems unreasonable.

However, I think the Venutian case highlights a benefit of the current account, not a defect. Because the account here is to make sense of how norms are related to their objects, and part of the motivation underlying the account is to balance the demand for a unified account of norms with the reasonable thought that norms shouldn't be imperialistically applied. The anti-imperialism is motivated by the thought that norms license evaluation of an individual, and should thus be about the properties, specifically the goodness or badness, of the individual being evaluated, not the properties or interests of the individual doing the evaluating. Norms, after all, are what allow us to consider something as better or worse as the thing it is, not for some use or function external to it.

So, we should extend our norm only to things which have the functions that support them. The trouble here is that the very explanation that allowed us to explain why Swamp-judgments of certain kinds still held—the presence of beliefs and desires independent of function—permits them to arise by various functions as well, thus allowing, in principle, some capacities that seem identical to our own to have functions that differ from our own.

These Venutians wouldn't have the functions we have, because they have etiologies that support divergent functions, and so their theoretical norms are different than ours. Those who hold this against me are begging the question against me. This isn't a consequence of my view, this is my view. Those who hold that it is impossible that different minded beings must all be governed by the same norms need to reject a central tenent motivating the view under consideration—that things are governed by norms supported by their own nature. Things of different natures run the chance of being governed by different norms. That things with different natures can share similar working mechanisms (in this case: mental capacities) is no more surprising than that machines that have different functions can share similar mechanisms.

In order for the Venutian worry to be detrimental to my position, an opponent must reject a central motivation of this project. Pointing out that my view entails that individuals with identical actual traits can be governed by different norms is pointing out a fundamental feature of my view. This is a *modus ponens* that cannot be *tolensed* without further argument on the part of the objector.

4.4 CONCLUSION

In this chapter we've seen several things. In 4.2, I made room for unity of the grounding of norms in their objects by first motivating unity as desirable in the domain of norms (4.2.1) and then seeing how unity might be had despite superficial linguistic differences in normative domains (4.2.2 and 4.2.3). In 4.2.4, I argued that the objects of theoretical norms should be considered primarily mental processes or other mental capacities, rather than people or reasoners. In 4.3, I then considered several conceptual objections to locating norms in

etiological proper functions. First (4.3.1), that we want to apply norms to things that might lack etiology (Swampman); second (4.3.2), that norms might be barred to things due to their different etiologies but which seem to share the same function (imagined Martians), and last (4.3.3), that norms might differ between groups having identical capacities because of divergent etiological functional (humans and imagined Venutians).

5.0 EMPIRICAL THREATS

5.1 INTRODUCTION

In Chapter Four, I motivated the unified account of normative explanation the metanormative picture on offer provides and defended it from several *a priori* objections stemming from its reliance on etiological function. In this chapter, I address another source of threat to the grounding of norms in etiological function—empirical results from contemporary psychology. Section 4.3.3 discussed the worry that possible beings with mental capacities that have identical current properties to ours might, nonetheless, have etiologies accounting for different functions that support different theoretical norms. There I accepted this possibility as a benefit of the current view—that it is anti-Imperialistic about the norms that govern individual kinds. This anti-imperialism is gained at the cost of epistemic certainty with regards to the appropriateness of the norms we apply. On this view, to know what the norms governing some individual are, you must know certain things about its nature and relevant facts about the nature is often removed from us by time.

The threats considered in this chapter all threaten to undermine the current metanormative account by raising the possibility that we ourselves are the hypothesized Venutians from 4.3.3. It threatens to put tension on the anti-imperialist by showing that we risk giving up too much. What if *we ourselves* have functions that either support counter-intuitive norms or lack the necessary functions to support our intuitive norms? The theory seems to be in a vulnerable position of being empirically falsifiable, and some current theorists and evolutionary psychologists think that the very accounts that would undermine our faith in the metanormative theory are likely.These results generate two basic worries for the metanormative constitutivist. The first worry is that at least some of our intuitive theoretical norms

aren't supported by an etiological function; this is a worry that the functions available to us are either too weak to do the job the metanormative theorist wants or requires a dramatic revision of our accepted theoretical norms. The second worry is that etiological functions can be found to support norms contrary to our intuitive theoretical norms; this is a worry that the functions available to us might be too strong and entail norms the metanormative theorist would want to reject. Both of these worries, then, push counterintuitive norms on the metanormative theorist who ties norms to etiological function. They both push us to make a choice: the norms we know and love or the norms we can get from etiological function. The purpose of this chapter is to come down on the side of "norms we can get from etiological function," while showing that the most intuitive norms that trouble those who raise the first worry are going to be supported on any plausible etiologically functional story and that the most troubling norms in the second worry are unsupported by etiologically functional stories.

I first canvas in 5.2, several of the more commonly discussed experimental results and some variants that are taken to shed light on what kinds of evolutionary stories might be the right ones to account for our abilities (or lack thereof). (5.2.1 Wason Selection Tasks; 5.2.2 Conjunction Fallacies; 5.2.3 Base-Rate Neglect; 5.2.4 Possibly Adaptive Delusional Beliefs.) Readers who are familiar with this literature may skip this survey at no cost to their ability to follow the argumentative sections that follow. In 5.3, I address two variants of the first worry that threaten to under-determine our application of normative evaluation and show that neither threatens to undermine our ability to negatively assess under-performance. In 5.3.1, I address the first variant, Subject-Based Threats, which turn on our inferential abilities being strongly influenced by the subject-matter of our inferences. In 5.3.2, I address the second variant, Input-Based Threats, which exploit the fact that the inferential abilities of experimental subjects seem to be strongly affected by the presentation method used in studies. Both threats make it seem likely that the mental capacities we take to be governed by our intuitive theoretical norms do not have functions that could support those norms. In 5.4, I respond to an objection that not only do we lack functions that could do the work to support our intuitive norms, we have functions that actively support counter-norms. In 5.4.1 I discuss psychological results that might be taken to show that not only are our mental

capacities not functional in ways that would support our intuitive norms, but that they are functional in ways that would support norms that are obviously irrational. These worries each threaten a disconnect between our intuitive theoretical norms and the norms that might be derivable from the functions of the things we assess according to those norms. I cannot here argue that for every intuitive theoretical norm there will be a proper function to support it or that for every proper function, it will only support our intuitive norms.[1] However, I will argue that the current psychological results canvassed above should not worry the aspiring metanormative constitutivist who is committed to defending intuitive rational norms. In each case, I argue that none of these results undermine our ability to derive intuitive norms from etiological functions.

5.2 PSYCHOLOGICAL RESULTS

In addition to worries arising from counter-factual and other *a priori* considerations, the constitutivist account developed here is also liable to raise empirical worries about whether our actual mental capacities have functions that support our intuitive theoretical norms. Many results in experimental psychology from the last thirty years demonstrate that human beings are characteristically bad at a range of cognitive tasks. Evolutionary psychologists have recently begun to give explanations of these results that might be taken to undermine our intuitive grasp on the norms governing these evolved capacities. Of course, being bad at certain tasks is not itself reason to think that an individual was not selected to perform them (remind yourself of a poorly designed artifact of your choice). Even where selection pressure is very high, what is statistically average for individual kind-members does not always correlate in the least to what it was selected for (think: mayflies). However, the evolutionary explanations of some of these systematic performance failures take them to be adaptive in ways that might be threatening to the constitutivist project.

Included here are four types of studies in the psychology literature that might be threat-

[1] More importantly, I think it is a mistake to want such an argument. Such an argument is one an imperialist about norms would demand. I reject such demands.

ening to this metanormative account. Each have results that show that we aren't very good, statistically speaking, at one of three sorts of things: selecting what kind of information we need to assess certain logically structured claims, reasoning according to probabilities, and forming accurate or justified beliefs about certain types of subject matter.

5.2.1 Wason Selection Tasks

In the experiments known as Wason Selection Tasks, subjects are shown four cards, each with a number on one side, and a letter on the other. Two cards are placed on a table with their letter-side up—a C and an E—and two cards are placed number-side up—a 5 and a 4. Subjects are asked which cards they would need to turn over to verify that the following claim is true: "If a card has a vowel on one side, then it has an odd number on the other side." (Variants use other phrases of the form 'if P then Q,')

Repeated studies have shown that people are surprisingly bad at selecting the correct cards to turn over to confirm the truth of the conditional claim. (The cards showing an E and 4 must be turned over—or whichever cards show a P or a non-Q.) Most subjects correctly indicate that E must be turned over, but many also claim (falsely) that the 5 must be turned over, even though regardless of what was on the other side the claim could not be falsified by that information. Most subjects also fail to indicate that the card displaying a 4 needs to be turned over, despite the fact that if it has a vowel on the other side, it would falsify the claim. These studies show that people are very bad (and predictably so) at certain types of conditional reasoning. When asked what sorts of information is needed to confirm conditional claims, subjects in these studies are terrible at selecting all and only the relevant information.

Selection task studies were first conducted by Peter Wason in 1966. Since then, variations on the original study have been conducted and shown that success rates among subjects seems to dramatically differ depending on the presentation method in the study.[2] Griggs and Cox (1982) conducted selection tasks where they replaced the vowels with "drinking beer" and non-vowels with "drinking coke" and the odd numbers with "25 years old" and

[2]See Koehler (1996) and Sperber, Cara, and Girotto, (1995) for discussions of variant explanations of Wason selection tasks.

even numbers with "16 years old." They told the following story to the subjects:

> In its crackdown against drunk drivers, Massachusetts law enforcement officials are revoking liquor licenses left and right. You are a bouncer in a Boston bar, and you'll lose your job unless you enforce the following law: If a person is drinking beer, then he must be over 20 years old.

The subjects are then asked which of the cards they will need to turn over to ensure that everyone drinking beer is at least 20 years old.

In this case, the change in the presentation of the test corresponded with a dramatic increase in performance. Where 25% of subjects were able to correctly pick out the E and 4 as the cards to turn over, 75% of subjects picked out the cards with "is drinking beer" and "is 16 years old" as the cards they needed to turn over to confirm the conditional claim. So, presentation appears to have a dramatic effect on whether we're any good at selecting all and only relevant information in these cases. A contender for an explanation of this disparity in performance is the Cheater Detection Hypothesis: we're selected to pick out cheaters, not solve logic problems.[3]

The Cheater Detection Hypothesis postulates that success rates rise when you change the presentation of the choices (and add a cheater-mechanism "engaging" back-story) because changing the presentation permits us to utilize a module that was selected because it enabled us to pick out cheaters, rather than to reason conditionally. If we were merely selected to solve logic problems, we'd be just as good at solving logically analogous problems without the cheater content.[4]

5.2.2 Conjunction Fallacies

In addition to reasoning with conditionals, we're also statistically bad at reasoning with probabilities and conjunctions. (In)famously, Tversky & Kahneman (1982) presented subjects

[3]For more on the Cheater Detection Hypothesis, see Trivers (1971) and Cosmides and Tooby (1992).

[4]I'm not going to address the literature questioning the Cheater Detection Hypothesis's interpretation of the role of subject matter in accounting for the differences in performance, because it is only on this interpretation that these results seem to cause trouble for my view. But for interested readers, see Carlisle and Shafir "Questioning the cheater-detection hypothesis: New studies with the selection task." and Brown, W. M., & Moore, C. (2000). "Is prospective altruist-detection an evolved solution to the adaptive problem of subtle cheating in cooperative ventures?" and Sperber, D., & Girotto, V. (2002a). "Does the selection task detect cheater-detection?"

with the following scenario about a woman named Linda:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.
> (a) Linda is a teacher in an elementary school.
> (b) Linda works in a bookstore and takes yoga classes.
> (c) Linda is active in the feminist movement.
> (d) Linda is a psychiatric social worker.
> (e) Linda is a member of the League of Women Voters.
> (f) Linda is a bankteller.
> (g) Linda is an insurance sales person.
> (h) Linda is a bankteller and is active in the feminist movement.

For any two conjuncts, the probability of the individual conjuncts must be higher (or at least equal to) the probability of their conjunction. The probability of (h) cannot be higher than the probability of (f). Yet compound events are regularly judged more probable than either of the components. Of those subjects with no background in statistics, 89% judge that (h) is more probable than (f). Of those subjects in graduate programs in business school, 85% make the same mistake. This study and others like it show that we (sometimes) underestimate the probability of individual conjuncts relative to our estimation of the probability of their conjunction. We're prone to the Conjunction Fallacy.

Similar studies that replace the probabilities with frequency reports have a surprising result: the rate at which people rank individual conjuncts as less probable than their conjunction is dramatically reduced. The frequentist presentation was:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
> There are 200 people who fit the description above. How many of them are: bank tellers? bank tellers and active in the feminist movement?
> ...

Fiedler (1988) reported that only 22% of subjects judged that there would be more feminist bank tellers than bank tellers when the phrasing listed above was used. When the test was run again, Gigerenzer (1994) reported that only 13% of subjects made the conjunction fallacy when asked to estimate only three things: the number of women out of 200 who were bank tellers, feminist bank tellers, and feminists.

Again, evolutionary psychologists have an account: they think that the low scores are the result of the foreignness of probabilities to our evolutionary ancestors. Cosmides and Tooby (1996) term it the Frequentist Hypothesis: Some of our reasoning mechanisms do deal with probability, but they are designed to take frequency data as input and output frequency judgments. So, the difference in performance levels on this account is explained by our being bad at probability calculations because either we are not evolved to make them at all or the evolutionary mechanism that makes them is worse than the one that makes frequency calculations.

5.2.3 Base-rate Neglect

Casscells et. al. (1978) presented the following question to a group of students and faculty at Harvard Medical School.

> If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually have the disease, assuming that you know nothing about the person's symptoms or signs?

The correct answer is 2%, but only eighteen percent of the audience answered close to that. Forty-five percent say that the answer was 95%, which is roughly the answer one would get when neglecting to consider the base-rate of the prevalence of the disease in the population. Researchers concluded that the faculty and students were failing to take base-rates into account when determining the actual likelihood of illness in the patient who tested positive. This sort of result has been frequently reproduced and seems to support the more general conclusion that people are generally bad at reasoning about probabilities given multiple probabilistic inputs.

As in conjunction probability, subjects' abilities to consider multiple inputs increases when the information is given as frequency, rather than probability data. This has also been explained by evolutionary psychologists by the Frequentist Hypothesis. When asking the number of people out of a hundred who would test positive for the disease given the base rate (given in frequency) and the false positive rate (also in frequency), 76% of respondents answered correctly. This is a drastic reversal, and might be taken to show that we are selected

to reason about frequencies, not probabilities. Together with the Conjunction Fallacy data, this seems to show that perhaps we were not selected to reason probabilistically at all.

5.2.4 Delusional Beliefs

Psychology literature discusses both statistically abnormal and normal delusions. The former include delusions that are the subject of psychological treatment and involve well-researched classes of delusions like Cotard Delusions (which manifest with persistent entrenched beliefs that one is dead) or Capgras Delusions (which manifest with persistent entrenched beliefs that a loved one has been replaced by an impostor). These, however, are not going to be relevant for critics of the functional account, because they are clearly not adaptive. They are not plausibly cases of reasoning that are candidates for adaptive misbeliefs, the term used by Dennett and McKay to pick out beliefs resulting from mental capacities selected by evolution to produce misbeliefs.[5] Not only are these delusions statistically anomalous, but they seem to have no adaptive benefit, present or past, that could explain why the process that produces them could have been selected *to* produce them.

Normal delusions, however, provide examples where we might find cause for concern. Positive delusions might pose the best chance of being adaptive and thereby threatening to the metanormative constitutivist.[6] There is a lot of research chronicling people's consistent tendency to over-rate themselves and loved ones in almost any area viewed to be positive and, correspondingly, to under-rate the prevalence of negatively-viewed traits.[7] Gagné and Lydon (2004) demonstrated, for example, that 95% of people judge their partners more positively than they judge the average person. Wegner & Fowers (2008) showed that 86% of people judged their children to more strongly manifest positive attributes than the average child while 82% of people judged their children to less strongly manifest negative attributes than the average child. Most everyone, it seems, thinks their loved-ones are above-average.

Interestingly, people are pretty consistent in these kinds of misbeliefs and aren't very

[5] McKay and Dennett "The Evolution of Misbelief"

[6] For a discussion of misbeliefs and the possibility of adaptive misbelief that settles on positive delusions, along with a helpful discussion of the state of the literature in this area, see McKay & Dennett (2009).

[7] For a discussion of general narcissistic bias, see Akins (1996), for discussion of what is known as better-than-average-effect, see Alicke (1985).

good at correcting for them, even when they're given compelling incentives to do so. This kind of over-assessment of ourselves and loved others is often tied to our tendency to be over-confident in our own abilities.

In one study, people are given the following questions:

In each of the following pairs, which city has more inhabitants?
(a) Las Vegas (b) Miami
(a) Sydney (b) Melbourne
(a) Hyderabad (b) Islamabad
(a) Bonn (b) Heidelberg

In each of the following pairs, which historical events happened first?
(a) Signing of the Magna Carta (b) Birth of Mohammed
(a) Death of Napoleon (b) Louisiana Purchase
(a) Lincoln's assassination (b) Birth of Queen Victoria[8]

After each answer the subject was asked to report their confidence in their choice. Surprisingly, there was a robust tendency to be over-confident on one's answers, given hard enough questions. People seemed to be about 20% overconfident (e.g. By reporting 80% confidence, when only answering 60% of questions correctly). This phenomenon seems unaffected by position; it has been found in undergrads, grad students, physicians, and even CIA analysts.[9] Moreover, the manifested overconfidence is very resistant to incentives for correction (e.g. even when offered rewards for accurate predictions). Even if you offer people rather large sums of money for an accurate appraisal of their performance, they still exhibit over-confidence. So, it seems to not be the result of subjects wanting to be viewed as more intelligent or confident.[10]

[8] Test yourself! The answers to all population questions is A and the answers to all chronology questions is B.

[9] See Lichtenstein, Fischoff, and Phillips (1982.)

[10] Another type of positive illusion, discussed in McKay & Dennett (2009) involves positive illusions about health. These are not, like the abnormal results I dismissed above, psychologically troubling, but they are not common like other over-confident false beliefs. There is a significant literature supporting the hypothesis that having unrealistic expectations about one's health-outcomes is highly correlated to improved health-outcomes. For example, there is research that shows that unrealistic expectations about everything from how likely one is to fight an HIV infection to whether one will survive breast cancer track success rates at these activities. For a discussion of the studies of HIV infected men and their survival expectations, see Reed et al. (1999).

5.3 UNDER-DETERMINATION THREATS

There are two related ways that the results in 5.2, can be explained by an etiological function such that it would not support our intuitive norms; I distinguish them by the names Subject-Based Threats and Input-Based Threats. Subject-Based Threats arise from evolutionary explanations of under-performance in specific subject-matters, which claim our mental capacities evolved to reason about a particular subject matter, rather than reason in general. This is a worry that could be raised by the Wason selection test results. Input-Based Threats arise from evolutionary explanations of under-performance linked to the mode of presentation, which claim that our mental capacities were evolved to have a particular sort of input. This worry is raised by possible evolutionary explanations of base-rate neglect and conditional probability. Both worries threaten to restrict the function of the selected mechanism in a way that could not account for all of our intuitive theoretical norms.

5.3.1 Subject-Based Threats

Wason selection tasks and related tests of conditional reasoning skills show that we're predictably bad at certain logic-based conditional reasoning tests, but not nearly as addled when doing similarly structured conditional reasoning tests with specific subject matter. Evolutionary psychologists stipulate that our mental capacities are the result of long-past selection pressures our evolutionary ancestors faced in the Pleistocene. This would allow us to account for the functions of the selected-for capacities by appeal to what adaptive problems our evolutionary ancestors might have faced. But the discrepancy between our performance on differing subject matters could indicate a threat to the metanormative constitutivist's reliance on etiological function.

If our evolutionary ancestors' mental capacities were selected to reason about one specific subject, and we are only able to generally reason as a derivative residual ability, then it would seem that those capacities do not have the function of general reasoning. If our ancestors passed those selected-for capacities to us, poor performance with conditionals could be explained because of a difference in the subject matter that the capacities were

selected for. But this explanation would lead us to conclude that failure to reason well with non-selected-for subject matter would not be irrational because the failure with the non-selected-for subject would not show that we were bad at performing the function of the relevant cognitive process.

The possibility that we have one conditional reasoning capacity selected for a specific cheater detection function is intuitively threatening. But before we consider just how threatening, it's important to note that this possibility is not viewed as the most plausible explanation of the subject-variant results of Wason Selection Tasks. The most commonly accepted explanation of the differing results is not that we have one general capacity that was selected for its ability to reason about cheaters, but that we have two (or more) capacities, (including) one that was selected because it enabled reasoning about cheaters and one that was selected for its general reasoning abilities. This dual (or multi) process account explains the discrepancy in performance ability as the result of different mechanisms being utilized for the different subject matters.[11] The subject of cheaters is thought to trigger what is known as a system one capacity (quick, automatic), while the general logic puzzle version of the task requires using a (slow, deliberative) system two capacity. This explanation would not pose a threat to the metanormative constitutivist because the two (or more) capacities could each have norms accounted for by different etiological functions. Such an account would just postulate (at least) two capacities, one that is not very effective, but which allows us to solve regular Wason selection-task type problems (but not with high frequency) and one that is much better at solving cheater-detection variants.[12]

So, the challenge of subject-based threats is a hypothetical one: *if* it were the case that we have one evolved conditional reasoning capacity, *and* it was originally selected because of its ability to detect cheaters, then (and *only then*) it wouldn't be irrational for it to be bad at general reasoning. Put aside for the moment whether the antecedents are true, it's not clear that the inference goes through, even given the truth of the evolutionary psychologists' account. The conclusion is only reached given one of two assumptions that I will reject in

[11] See Frankish (2010) for an overview of the current literature on dual-process views of reasoning.

[12] There might be a residual worry that this might make a priority between the different system norms, so that failing to be able to catch cheaters might be *more* irrational than failing to be able to perform certain system two functions. I doubt that this is a real worry, but insofar as it is, I think the worry will be about whether we can have conflicting normative systems. I'll begin to address this worry later.

the remainder of this section. First, it would follow if once a form is selected because of its performance of some function, another function could not be gained without a change in form or active selection of that form for a new function. I will argue below that this stagnant view of function doesn't hold in either artifactual or biological cases because of the role adaptive problems play in function acquisition for adaptive functional kinds. Second, even given this general claim, the concerning conclusion would still follow if our reasoning capacities weren't the sorts of things that could have additional functions in the way the first response provides. I will argue that there is an adaptive problem that would account for the more general function to have been acquired by the form originally selected for its cheater catching capacity.

So, assume that we have one conditional reasoning capacity that was originally selected for and distributed amongst our evolutionary ancestors because it helped them catch cheaters, and thus solved an adaptive problem. This, at least, commits us to the position that our conditional reasoning capacity has the function of catching cheaters. It doesn't show that the capacity lacks other functions. This only follows if you accept a stagnant conception of function acquisition, according to which once a form or trait has acquired a function through a creation process, it cannot gain others. Given the discussion in Chapter Two, this position might seem plausible, but the account there is sympathetic with a more dynamic account of function acquisition. In cases of both artifacts and biological kinds, forms can gain functions without being significantly altered. I'll quickly give an example of each to illustrate the phenomenon and then explain the general principle behind function acquisition without new creation processes.

The artifact case is about Otherworldly Knives (OWKs). Imagine we inhabit a world in which bread is the only substance that we desire to cut and we have designed an artifact to cut and slice bread. We've designed bread knives, which have the proper function of bread cutting. Our OWKs are bread knives, but they would be pretty good at cutting other things, if they were around to be cut. Our R&D lab has also invented a space shuttle, and we venture to Earth, where among many other things to cut, we encounter fish. Our OWKs turn out to be pretty good at cutting fish, though not as good as at cutting bread. Are our knives bad because of this? Our account says: no. Our OWKs are not fish-cutters, they're

bread-cutters. It is not the use an artifact is put to that determines its function.

But, something can happen to our OWKs that could give them the function of fish-cutting. If we were to develop a taste for fish, and a need to cut them, we could solve this adaptive/engineering problem in two ways. We could manufacture new knives, fish knives, or we could leave well enough alone with our OWKs. If we're not very picky, and are happy to have one tool that is great at bread cutting and pretty good at fish cutting, once we have begun to manufacture OWKs in the original form to serve the dual purpose, that form gains the dual function. And, if it turns out that we realize we want something that's going to be pretty good at general cutting, and produce them for this purpose, we then would have a general cutting knife. This can all be true even if the original design is unchanged *and* the OWK is still *best* at bread cutting.

So, it seems that in the case of artifacts, it's perfectly plausible that a form that was originally selected to perform a subject-specific function can gain more general functions without changing forms. In the OWK case, what once was a specific-use item gained the function of a general-use item via selection processes. The form that was selected was the exact form that existed already, but that's the way selection sometimes goes. The continuation of instances of the form is explainable because of the fish-cutting as well as the bread cutting function. The form did not change, but we can see how the function did.

The biological case might seem more challenging because we cannot appeal to user need and selection or appeal to a counter-factual scenario where the form did change to meet the new adaptive problem. But it is widely recognized that in cases of biological kinds, changes in function can occur without changes in form. These kinds of functions are often referred to as exaptations, the use of existing adaptations for the solution to newly occurring adaptive problems. It is common to talk about forms being selected for functions as if evolution were an inventor who thought up a clever solution to an engineering problem, but we know that isn't the case. The process of evolution works with existing traits and retains those that provide solutions to adaptive problems via reproductive selection pressures. In cases where the trait is already distributed throughout the genetic pool because of a prior selection event, that the trait or form was previously selected for another function does not preclude it from being retained for another purpose and thus for gaining another function.

An example of such an existing functional trait that was selected because it solved a new adaptive problem is feathers. Biologists believe that feathers were originally selected and spread throughout the bird population because of their insulating properties.[13] But when birds became otherwise capable of flight, and thus needed buoyancy and lift, the same form that was so good at insulating was also able to provide flight. It is now true both that feathers have the form they have because it was selected because it insulated and that they have the form they have because it was selected because it provided flight. That the trait was existing in the population when it was selected because it provided flight is no more relevant to the selection for flight than that the trait existed in one member of the population when it was originally selected for its insulation. Its ability to provide flight accounted for its continued appearance in the population, even though it does not account for its dispersal throughout the population, because it solved an adaptive problem.

The appearance of adaptive problems and the traits that can serve as their solution do not generally coincide. When feathers were originally introduced (as a mutation) into the population, it need not have been the case that the adaptive problem of insulation was present. There is often an evolutionary lag time in one direction or another—sometimes traits occur in a population before an adaptive problem they could solve, and sometimes adaptive problems occur before a mutation arises that could address the problem. Cases where forms with pre-existing functional forms gain new functions are cases in which there is no evolutionary lag time. But that there is no evolutionary lag time, or that there is another function, doesn't prevent a new function from being gained when an adaptive problem is addressed by a trait. We should, then, reject a stagnant conception of function acquisition in favor of a more dynamic one. It is possible, both in artifactual and biological kinds, for forms to gain new functions after their original etiological creation endowed them with a first function.

Despite it being possible for existing forms to gain new functions, in order to show that

[13]It has been impressed on me by Edouard Machery that it is highly unlikely that there was not a change in the form of feathers throughout their evolutionary history. I take this point, but don't think this is a particularly problematic for the view here. At issue here is whether mental capacities can have exaptation functions, not whether exaptation functions ever involve change of form. Everyone can likely agree that both feathers and mental capacities have changed in form since their original selection events. The question is whether they can have gained additional functions without those forms being selected for.

the threatening evolutionary story of the cheater detection conditional reasoning mechanism is not a threat, the mechanism originally selected for cheater detection must plausibly have gained a general reasoning function.[14] Both the above accounts worked by showing that there was some new adaptive problem that the already existing form addressed, such that, if it had not addressed it, individuals with that form would have been selected against for that reason (feathers) or another form would have been introduced (OWKs). To show that it is plausible that the original cheater detection capacity gained a general reasoning function, we need to show that it likely addressed an adaptive problem that played an evolutionary role in the survival of our evolutionary ancestors. So, if there are ancestors who continued to be reproductively successful because of the general reasoning abilities provided by the existing specific cheater capacity, then it will have general reasoning as a function as well.

The adaptive problem can't be additional specific subject matter (this would just compound the number of specific functions, not lead to a general one). The adaptive problem is evolutionary lag time itself. It is highly likely that the general ability of conditional reasoning capacity solves a real adaptive problem with evolutionary lag time. Having evolution wait on genetic mutations is a slow and taxing process.[15] It requires waiting on mutations and hoping the individuals that manifest the mutations are otherwise reproductively fit, and all the while the adaptive problem is hindering the flourishing of the population. If there are new subject matters to reason about all the time (today cheaters, tomorrow the migration patterns of wildebeest) having a general capacity that could be counted upon to be there regardless of the subject would avoid the lengthy wait on evolution to find a mutation to select into the genetic pool. If the cheater detection mechanism was generally good at reasoning, it is likely that it would solve the adaptive problem of evolutionary lag time in the

[14] I am here assuming, to take the strongest version of this challenge, that the form of our conditional reasoning mechanism has remained unchanged since it was selected into the population of our evolutionary ancestors. This strikes me as highly improbable, and if, as seems more likely, versions of the mechanism that were better at general reasoning were selected for over versions less good for general reasoning, this makes my position even stronger.

[15] It is worth noting that domain-general reasoning capacities are not the only method that could address the adaptive problem of evolutionary lag time. There are non-rational learning capacities and cultural transmission of knowledge is another. However, I am not claiming here that it is the only thing that *could* address the adaptive problem of evolutionary lag time. I am rather claiming that it has addressed the adaptive problem of evolutionary lag time. I want to thank Edouard Machery for bringing this issue to my attention.

domain of reasoning.[16] Evolutionary lag times are *themselves* adaptive problems that a general reasoning capacity seems primed to deal with. We need not wait for a genetic mutation to give us an additional specific reasoning capacity, because we have a general one. That is, it's not merely a pleasant side-effect of the cheater catcher mechanism that we can reason generally. It might have been originally, but it has solved a very real adaptive problem for us it has made us able to reason about our needs rather than wait for the slow process of evolution to do this work for us.

This account makes our reasoning capacities and their associated functions no more mysterious than the dual functions of feathers, first selected for insulation and which later gained the function of providing lift for flight. No change in form must be postulated to account for the additional function, though an additional adaptive problem must be found that the form would have been selected for, that can then be a further explanation of the continued presence of the form. In the case at issue, independently of the ability to catch cheaters, the form of the reasoning mechanism would have been retained because it served the purpose of reasoning about new subjects as they arose instead of waiting for evolution to provide piece-meal mechanisms or having some other solution to evolutionary lag time.

Subject Based Threats are thus no threat at all. It is possible for functions that are subject-specific to become subject-general without being accounted for by a selected-for change in the form. Functions of a kind need not be static, even when the form might be. These experimental results, then, don't show that our selected-for capacities lack general conditional reasoning functions. Grant the worst case evolutionary explanation: that the capacity we think of as a general conditional reasoning capacity was originally selected in order to catch cheaters. This can explain why we are still better at catching cheaters than at performing subject-less conditional reasoning tasks. What it does not do is prohibit the thing that was originally selected for its cheater catching abilities to be retained and gain

[16] One might wonder, as has been pressed to me by Edouard Machary and others, if we are evolved to be generally good at reasoning, why are we so demonstrably bad at some general reasoning tasks? I think the answer to this is that we're *generally* good, and generally good given the Normal conditions of the selection environment. One might also wonder, then, why general reasoning capacities aren't more widespread among animals. I think this seems best explained by different ways of dealing with these adaptive problems being selected for. Perhaps we addressed this adaptive problem with a general-capacity reasoning mechanism and non-reasoning forms of learning, while other animals did the same with non-reasoning forms of learning alone.

the function of reasoning more generally. The subject-based threats are thus not a threat to the metanormative constitutivist.

5.3.2 Input-Based Threats

Subject Based Threats turn on evolutionary accounts that postulate what we take to be general mechanisms were originally selected to address a specific adaptive problem. This threatens to make addressing that specific problem the sole function of the mechanism, thus undermining our ability to apply more general norms to it. The resolution just presented turned on being able to account for how a mechanism or form in general could gain further functions by later serving other, possibly more general, adaptive problems.

Input-Based Threats have a different valence. They turn on evolutionary accounts that postulate specific mechanisms selected to serve a seemingly general adaptive problem. This threatens to make any non-selected-for specific or general input immune to evaluation according to the norms. The resolution to the subject-based threats is thus not applicable; the problem isn't that the form needs to gain a new function, but that we need to account for how the function and norms need not be constrained by a specific aspect of a selected-for form.

Evolutionary accounts that lead to the Input-Based Threats are sometimes given to account for the results of studies like those in the Conjunction Fallacy and Base Rate Neglect sections above. In these studies, performance varies depending on how the problems were posed.[17]But instead of subject-based differences, base-rate neglect and conjunctive probability assessment performance seemed affected by the way that the information was presented. In these studies, the ability of subjects to accurately process content depended on whether it was presented as probability data or frequency data.[18] The difference in abilities has led

[17]It's important to remember that we *can* solve both types of inputs. We're much better at frequency assessment than probability, but we can make probability assessments, and at a much better rate than chance. Contrarily, it might seem puzzling how poor we can be at reasoning tasks if we are evolved to deal with them. Edouard Machery has pressed to me that while natural selection does not optimize, it at least always moves towards optimality. I think this is right, but it's important to see that adaptive pressures are not always pressing in the same direction.

[18]It's important to note here that I am not taking, nor do I need to take, a stance on the relationship between probabilities and frequencies. Frequentist views of probability hold that the two are identical, that probabilities are frequencies taken over finite or infinite instances. So, on this view, the probability of a

some evolutionary psychologists to hypothesize that we are (statistically) better at solving problems when stated in frequencies than in probabilities because the mechanism that was selected because it dealt with these sorts of problems was selected because it processed frequency inputs, not probability inputs.[19] Evolutionary psychologists have a fairly plausible story for how the selection might have worked: in the Pleistocene, we were presented with frequency information, not probability information, and so we developed mental processes that were more adept at processing frequency data to make judgments about likelihood.[20] The account assumes that there are (at least) two ways to reason about uncertain events or states of affairs: by considering probabilities and by considering frequencies. The risk is that we're just designed in a way that lets us off the hook, evaluatively speaking, for some of our low performance on probability-based reasoning tasks. Because the mechanism wasn't selected to take probability inputs, it can't be negatively assessed for failures to accurately calculate probabilities.

The subject-based threats were answered by showing that it is possible in general (and likely in the relevant case) that mechanisms that once had functions relating to specific subject matter could have gained more general functions. The response to the input-based threats instead must show that there was a general adaptive problem that picked out a general function all along. The metanormative constitutivist needs to show that being selected to perform a function in a specific way does not entail that individuals with that selection history are not negatively assessable when they fail to perform the function in alternate ways.

---

fair coin landing heads is the frequency that it does so over a finite number of tosses or, alternately, over an infinite and possibly counterfactual set of tosses. There are well-known problems for both finite and hypothetical frequentist accounts. Hàjek's entry on Probability in the Stanford Encyclopedia of Philosophy has a nice overview of frequentism and its challenges. I only need the (I hope) independent claim that the same information can be presented as frequency or probability. It seems plausible to me that this can be the case, even if probabilities are actually different from frequencies. I only need the conclusion that frequency and probability-based mechanisms (if they are different) are governed by the same norms, not that frequency and probability are identical.

[19]See e.g. Gigerenzer et al (1999) and Todd & Gigerenzer (2000) which show that subjects' performances improve dramatically when simple frequencies are used to present probability data.

[20]I say that this is fairly plausible, but I'm actually not sure what it means to be confronted with frequency data rather than probability data. You can be confronted with many cases, but what makes these frequency-presentations? Frequencies and probabilities both seem like ways of representing the thing that is presented to you, and so what was presented to our evolutionary ancestors on the Pleistocene doesn't seem to be either frequencies or probabilities, but events.

This is, admittedly, a difficult task for the metanormative constitutivist. There are many cases where there are very different ways of performing a function that involve very different norms. In such a case, it is obvious that an individual selected to perform function Z by activity A should not be negatively assessed because it is unable to Z by B-ing. That a corkscrew is designed to open a bottle by removing the cork does not mean it is negatively assessable because it is unable to open a bottle by removing a bottle cap. In some cases, however, being designed to do Z by A-ing does open an individual to judgments for more general ways of Z-ing. There is a sense in which 'opening a bottle' does not fully describe a function, because there are so many ways in which bottles can be sealed, and being good at one doesn't always get you anyway towards being good at another.[21] Being designed to remove corks and being designed to remove bottle caps are two distinct functions. That both serve to open bottles is almost a distraction from this point. In contrast, being designed to remove corks by screwing into the cork and being designed to remove corks by wedging the cork out are two ways of being designed with the function of removing corks. And as it turns out, one is much much better than the other at the function. Being designed to do something badly is not generally a defense against accusations of badness. But at what point do we have two functions instead of one? This is a difficult problem that the input-based threat exploits.

The case at hand, that we're better at solving frequency equations than probability equations, seems to be one where it is unclear whether we have one or two functions. Though I think the argumentative burden is on the critic to show that the evolutionary account entails that there are two functions rather than one, I'll first try to motivate a positive case here for there being one function that has two ways of being achieved (e.g. removing a cork by pulling vs removing a cork by wedging) rather than two distinct functions (e.g. opening a

[21] There's an immediate worry here that this will lead to extremely thin functions, because we can always imagine a mechanism that does something well, but not some closely related thing well. So, why are opening a bottle by removing a cork and opening a bottle by removing a cap different functions while opening a large bottle and opening a small bottle do not seem to be? This is a worry raised by the appearance that what is being presented here is a principle for individuating functions on the basis of properties of things with those functions. This isn't the case. The principle being appealed to here is that there different *ways* of opening bottles, and these can be distinguished by recognizing them as different functions. There are not different *ways* of opening big and small bottles (except in extreme cases, in which case, these do seem like they will yield different functions). Thanks to Edouard Machery for pushing this point.

bottle by removing a cork vs opening a bottle by removing a cap). I'll then give two reasons that I think it's unlikely that the critic will be able to make such a case for two functions plausible. I'll first argue that such the evidence current taken to support such a divide is unlikely to be persuasive. Then I'll argue that the positive evidence is unlikely, given the kinds of evolutionary account that would be needed to support the dual function position.

The threat turns on there being two functions that support contradictory norms. But the evidence to support this possibility is that we are worse at reasoning with one way of presenting information (probabilities) than we are with another (frequency). So far, I see no reason to explain this discrepancy by appeal to the existence of two functions, and one mental capacity that was selected to perform one (reasoning with frequencies) while it was only derivatively able to perform the other (reasoning with probability input). That a capacity performs better with certain parameters than another does not entail that it was selected to perform the first, and it is no evidence that the parameters constitute different functions. I might, for instance, be bad at picking out red from brown beads, but excellent at picking out yellow from blue beads, because I am colorblind.[22]It does not follow from this that picking out red beads isn't a function of my eyes, while picking out yellow beads is. That something is bad at a certain aspect of a skill doesn't divide the skill into separate functions, only one of which it can be said to have. It also doesn't give us any indication that the responsible trait was selected to do only one of the (by hypothesis) two functions.

One should not rate the likelihood of a conjunction higher than the likelihood of any individual conjunct. This is a norm that people are better at reasoning in accordance with when the rating is asked for in frequency rather than probability terms. The results in cases like the Linda study show that for some reason we are able to process information more readily in frequency format than in probability format. This doesn't entail that there are two functions, or that the mental capacity responsible has the etiological function for reasoning with frequency but not probability.

[22]One might think that this illicitly appeals to a contrast between an individual and the population it belongs to and that, if all humans were colorblind, then it would be an indication that the function of vision was not to determine red from brown. I do not think this is the case. I think that if all humans were colorblind, this would not reveal that the function of vision were to only see some colors. As it turns out, humans are blind to some colors, and I think that our eyes are thereby worse at seeing. Other species that are able to see more of the color spectrum have better eyes.

In addition to the proposed evidence being too weak to support the conclusion that our mental capacities have been selected because they perform the function of frequency reasoning instead of the function of probability reasoning, I think there's reason to suspect such evidence is not forthcoming. The best evidence for such a two function account, where our mental capacities were selected for just one, would be if there were two sets of things to accommodate and that in our evolutionary history we were presented with just one. Moreover, if having a mental capacity to address this adaptive problem would account for why we were better at one than the other, this would support the input based objection. (Note, this is, wrongly, what the theorists appealed to take themselves to have made plausible.) However, there is no likely adaptive problem our evolutionary ancestors would have confronted that adaptation to frequency would have addressed that adaptation to probability would not have.

There is likely only one adaptive pressure in our evolutionary history: reasoning about uncertainty. Though we can present uncertainty as probabilistic or frequency data to test college students in experiments, our evolutionary ancestors did not get the data presented this way. Things in the Pleistocene (like most things today) were presented in neither probabilistic nor frequency data. We aren't presented with two different things when we're confronted with probabilities and frequencies. You can't see a frequency or find a probability in the wild. Both are ways of representation. If we had different capacities for dealing with these different presentations of the same thing, this would be easily accommodated by the metanormative constitutivist. Importantly, the evolutionary story we are presented with is not one where there are two capacities, one of which is better than the other. The evolutionary story under consideration has it that we were selected to have one capacity that performed a general function with exposure to one kind of input, while it also was able to deal with another kind of input, though less well. But it seems highly unlikely that there could have been one kind of input on the Pleistocene rather than another. So, there was no adaptive pressure to reason in one way or the other that was itself selected for in a way that could account for a more narrow understanding of the function. Unlike in the Subject-based threat, where the evolutionary story seemed plausible, and we had to show that it didn't cause problems for the metanormative account, here the evolutionary story doesn't seem

plausible at all.

For the moment, though, let's imagine you could be (and our evolutionary ancestors were) presented with frequency data, or what seems likely, that frequency questions more closely mimic the actual conditions of being presented with information like our evolutionary ancestors faced. If so, we're still left with the question of why the norms from the general function aren't constrained by the specific selected-for input. It will be useful here to introduce a distinction between being culpably bad and being non-culpably bad.[23] To be culpably bad requires having the proper function that licenses the norm you are violating and requires being selected to perform that function in the way you are bad at. So, there is a real sense in which *you* and not merely *your reproductively established family* is bad. Being non-culpably bad requires having the proper function, but having a form selected that is itself non-ideal. So, if the form that was selected is itself non-ideal, individuals that are resultantly in violation of certain functional norms are in that regard non-culpably bad. But they are still bad. If the form of a REF is deviated from by an individual, and resultantly the individual is in violation of a functional norm, the individual is thereby culpably bad. So, a corkscrew that has been bent from years of use might thereby be culpably bad, while a corkscrew that was created too small (because it is also a keyfob) might thereby be non-culpably bad. A corkscrew that cannot open a beer bottle is not thereby bad at all. The difference between culpable and non-culpable badness is whether the individual has deviated from the way they were created to perform their function or have the traits needed to perform their function as they were designed to.

This distinction allows us to make sense of a way that a mechanism could be functioning *as designed*, while still being in violation of norms that legitimately govern it on the metanormative constitutivist account. And this is a distinction that any view committed to the possibility of bad design must recognize. There are bad ways of designing artifacts; the IKEA knife and the Wüsthoff knife do not have different functions that they each optimally perform. They have the same function, and bad design accounts for why, in general, IKEA knives are so much worse than the Wüsthoff versions.

The evolutionary account of the conditional probability assessment results seems to show

[23] I pick this distinction up from Dennett and McKay, but it seems to be in common use.

that we are the IKEA knives of the reasoning world. The adaptive problem that the mechanism that deals with probability and frequency assessments solved was reasoning about the future. There's no reason to think that something about frequency itself was particularly adaptively beneficial; it is an accidental aspect of evolutionary history. Of course, accidents are important, and we can see that the prevalence of frequency presentation over probability presentation could account for *how* the mechanism was formed. But it doesn't account for a different function, and it doesn't allow us an escape from normative appraisal in situations with probability based information.

Our mental mechanism(s) that reason about the future and unknown likelihoods are, on this account, non-culpably bad at reasoning about probabilities. They weren't selected to take inputs as probabilities, but are none the less bad when they fail to reason correctly with these inputs because two conditions are met: first, they were selected to be reasoners about uncertainty, and two, reasoning about uncertainty requires reasoning well according to the norms about weighing likelihood. And facts about likelihood are equally capable of being presented as probabilities and frequencies.[24]The form of the mechanism was selected because it performed a general task that at the time of selection was achievable by more than one method. Importantly, though, the method selected doesn't seem to have been constrained by adaptive pressure.

We don't have to show that at the beginning the function was merely processing frequency data and then explain how the function of processing probability data was gained. The function was processing data about uncertainty, the mechanism was just selected to perform this function in a way that isn't as effective as possible. The resulting forms of defect are thus non-culpable,but still negatively assessable. Being selected to be pretty bad at performing a function is no defense against judgments of defect.[25]

The Input-Based Threat of the evolutionary psychologist is thus not a problem for the metanormative constitutivist. It merely shows that we might not have ideally designed mental mechanisms. This (sadly) is an obvious problem with evolutionary and artifactual

[24]This is not, strictly speaking, true. Some frequentists will deny single-case probabilities, but these cases are not going to be relevant to our question.

[25]It might seem counterintuitive to say that we are non-culpable when we, say, ignore the base rate. I think that we are just as non-culpable as are badly designed corkscrews. Of course, we can *learn* to be better at certain reasoning tasks, and perhaps we can be culpably bad, in that respect, when we do not do so.

creation: sometimes you get substandard gear to meet your functional needs.

5.4 ACTUAL FUNCTIONS YIELD WRONG NORMS

While some of the empirical results seem to lead us to functions that can't support our intuitive norms, some seem to lead us to functions that lead us to entirely counter-intuitive norms. The last concern, which I call Adaptive-Norm Violation threatens to give us norms that are actively counter to our intuitive theoretical norms. This sort of worry is raised by the adaptive misbelief literature. A misbelief is any non-justified belief, not necessarily false beliefs. In proposed cases of adaptive misbelief there is a tension between the intuitive norms we take to govern the functional kind, and the norms that we seem to actually be able to derive from the etiological function of the kind. These worries, if left unanswered, threaten to force us to look elsewhere for the norms governing evaluative kinds, undermining the metanormative constitutivist.

5.4.1 Adaptive Norm-Violation Threats

The delusional belief results are taken by many to be likely to lead to adaptive irrationality results.[26] While the earlier two concerns in section 5.3, addressed ways that our ability to negatively assess very common errors in reasoning might be undercut by evolutionary accounts, positive delusions threaten to pose a functionally adaptive yet irrational case. These cases provide functions that support counter-intuitive norms. McKay and Dennett write:

> In positive-illusions situations... the benefits accrue from misbelief directly—not merely from the systems that produce it... [S]uch doxastic departures from reality—such apparent limitations of veridicality—are not culpable but entirely forgivable: design *features,* even. These beliefs are "Normal" in the capitalized, Millikanian sense. In such situations, we claim, we have our best candidates for *evolved misbelief.*[27]

[26]Dennett and McKay canvas adaptive irrationality possibilities and endorse delusions as the most likely in their (2009).

[27]McKay & Dennett (2009) p. 507.

McKay and Dennett determine that positive illusions are likely to be adaptive, because they are prevalent and accord positive benefit. Of course, for them to be adaptive in a sense the metanormative constitutivist would be concerned by, the adaptiveness would need to be etiological, not merely currently beneficial. The cases McKay and Dennett consider are over-estimation of the positive attributes of one's romantic partner (by hypothesis, selected to: facilitate extended pair bonding), over-estimation of the positive attributes of one's children (by hypothesis, selected to: facilitate the care of otherwise annoying and taxing young), and positive illusions about one's likelihood to recover from illness or survive surgery (by hypothesis, selected to: interact with neuroendocrine system to instigate self-healing.)[28] If these misbeliefs are adaptive (i.e. they have the form they do for a selected-for effect) then it seems that we would have a case where negative normative assessment would be ruled out on the metanormative account.

Importantly, in order for this objection to be any threat at all, it must be the mental capacities or inferential practices that have non-standard functions. Though it is misbelief that McKay and Dennett consider adaptive, it is mental capacities or mechanisms or processes that must themselves be selected by evolutionary selection pressure. Beliefs aren't the sort of thing that can be selected for, evolutionarily, they are not genetically transmitted.[29] So, you can't select for false beliefs, only for more or less accurate belief-forming mechanisms/capacities. In the extreme, a belief-forming mechanism that is anti-truth-tracking could be selected for that reason, but it is still the mechanism that is selected-for, not the false beliefs.

The two previous threats involved cases where the evolutionary story seemed to make certain forms of theoretical irrationality outside the scope of evaluation. This threat poses to make the error itself selected for in a way that makes it the function of the form or mechanism in question. Instead of making us unable to apply norms to certain exercises of a mechanism, it threatens to force us to positively assess certain intuitive rational failings.

[28]For this last hypothesis, see Taylor et. al. (2000) and (2003). Note that this would be a particularly strange thing to have been selected *for* evolutionarily-speaking.

[29]McKay and Dennett recognize the distinction, and even go so far as to point out that it could only be the processes that are better, not the misbeliefs, "Even if they spring from adaptively biased misbelief-producing systems, therefore, individual misbeliefs about success are arguably more of a tolerable by-product than an adaptation." p 507 However, they still endorse the position that some misbeliefs could themselves be adaptive, not merely the systems that produce them.

We want to call over-rating your spouse and your own performance on multiple choice tests epistemically irrational, but if there is an evolutionary explanation that accounts for having the mechanism that leads to that over-ranking *in order to over-rank your spouse*, then it seems like the mechanism is a spouse over-ranker, and is functioning perfectly when it over-ranks the spouse.

We saw earlier, in section 5.4.1, that one form can have two functions. The evolutionary explanation accounting for the over-ranking function doesn't need to be undermined to show that we can negatively assess the over-ranking. What we need to show is first, that two functions can be had by the same form or trait that generate conflicting norms, and second, give an account that makes it likely that if we have one mechanism that has two functions, one to over-rank spouses and another to accurately represent things as they are. McKay and Dennett could be right that positive illusions are adaptive, but still be wrong that this is a genuine threat to our intuitive norms of theoretical rationality. For it to threaten our norms, the adaptive story would need to undermine our ability to assess the mechanism by the norms of rationality, not merely add additional, possibly counter-intuitive, norms.

Sometimes we assume that if x ought to do A this entails that x ought not do things that prevent doing A well. But this is unfortunately false. Of course, in normal circumstances, the implication holds. But if x is governed by two sets of norms, one qua A-er and one qua B-er, it's possible that the norms will conflict. Corkscrews remove corks from bottles. Keychains ought to be small enough to fit a pocket. But it could be (seems to be, judging by existing designs) that being small enough to fit in a pocket prevents needed features to remove corks from bottles efficiently. But a corkscrew keychain is still assessable by both sets of norms. It's just a bit of a tragic item—or a glorious item, depending on how frequently you need to open bottles on the run. There are any number of badly designed kitchen gadgets we could explore here, but suffice it to say that norms can conflict. It's not just whimsical kitchen implement designers that are the cause of these conflicting evaluative kind hybrids, either. There's no reason why the sort of evolutionary function-gaining story told in 5.4.1 couldn't be for an adaptive problem that could not, *in principle* be ideally solved by the trait in question. Say providing lift for flight couldn't be best performed by something that was insulating; there's no reason to think that this would prevent the same account of feathers

having both insulating and flight-providing functions. It's just that the feathers couldn't ideally satisfy both functions.

So, we need to demonstrate that we have an available negative appraisal for the spouse-over-ranker. This doesn't commit us to saying there's nothing they're doing right, after all; I don't think we should say that. It's good to think highly of your spouse, possibly even to over rate them. It just commits us giving an account of how one mechanism can be both for making sense of the world and for corrupting your perception of it. How those two things could both solve adaptive problems is pretty straight-forward: both would accord selection benefits on the individuals in a population who were able to solve those problems. Our reasoning capacities might sometimes be bad in ways that are beneficial for us enough to allow us to have, say, some misbeliefs permitted. The pair-bonding facilitated by thinking your wife is the most beautiful woman in the world, say, seems likely to cause fewer problems than the diminutive nature of the bad corkscrew. Both are misfires, all the same.

The pair-bonding that the misbelief is good *for* is external to the reasoning itself. Pair-bonding isn't a matter of mental capacities; it permits the organism that succeeds in establishing the bonded pair to more effectively raise young and propagate genes (or something), but it doesn't make the rational capacity better at the thing that rational capacity was selected for. It makes it worse by, by hypothesis, making it also something else: a pair-bonding facilitator. The goodness of the rational capacity is sacrificed for the good of some other end. Just like in the case of the keychain corkscrew. Sometimes we're willing to sacrifice good-making corkscrew features to make them highly portable.

It's consistent with the possible pair-bonding facilities of our rational capacities that we are thereby worse rational agents when we over-estimate the positive qualities of our spouse. It is, of course, possible for something that has two functions to have two functions that cannot be ideally realized simultaneously. Sometimes development of one function requires a sub-optimal form for the other. This does not hinder our ability to negatively assess our overly optimistic judgments of ourselves and others. We might be irrational in having them, but also be good spouses in virtue of having them. This seems like the right thing to say in both cases. Thinking your wife is more beautiful than any other woman seems both to make you a sub-optimal reasoner and a good spouse. Maybe in reasoning and partnering (like in

so much else) we can't have it all.

This only threatens negative assessment according to our theoretical norms if having the function of promoting pair-bonding somehow prevents the capacity of also having a general reasoning function. But I don't see any reason to think that it does. Unlike in the other evolutionary explanations in 5.4, there doesn't seem to be any claim here that the original selection of the mechanism was for the pair-bonding function. Absent such an account, showing that two conflicting functions are possible and indicating how that could be the case here should be sufficient to show that we can maintain our epistemic disapproval of the spousal over-ranking and other positive misbeliefs while accounting for them as possibly selected-for. Nothing here prevents the negative assessment according to the norms of theoretical reason. We are no more free of negative assessment because of our positive illusions than the mini-corkscrew is because of its design as a key-fob. Tragedy is not exculpatory.

5.4.2 Contradictions and Irrationality

Though one might be convinced by the account above that it is irrational to (say) over-rate one's spouse, it might still be open to an objector to wonder whether the theory at hand could be committed to it *also* being rational to over-rate the spouse.

This predicament arises from the combination of two factors. First, I have argued in 5.4.1 that it is possible that we have mental capacities that have two functions: to generally reason, and to (say) over-rate spouses. I argued there that this still allows us to negatively assess the over-rating, because it is in violation of the norms that follow from the function of generally reasoning. I likened this to a dual-function tool that is worse at one function in virtue of having the second. This is unproblematic in general, but combined with the particularly impoverished understanding of rationality that I introduced in Chapter Three, this could prove to raise an additional threat.

In Chapter Three, I said that theoretical rational is the evaluative term used to pick out those who satisfy to a high degree the norms that follow from the function of mental capacities. But no restriction was put on what those functions could be, and so here we have

a case where we are postulating that mental capacities could have functions like "over-rate your spouse, relative to the general population." It would be particularly bad if these were the *only* functions that mental capacities had, but it is still a hard pill to swallow that these sorts of mental functions could have a say in rationality. It will lead to two problems: first, the counter-intuitive result that it is rational to over-rate your spouse and second, the even more counter-intuitive result that it can be *both* rational *and* irrational to over-rate your spouse. It's counter-intuitive because of the seeming contradiction here. How can it be both rational and irrational to do the same thing? Shouldn't rationality be univocal?

I think there are two ways to go in responding to this challenge. First, to accept the seeming contradiction, but argue that it is not so difficult to imagine something being both rational and irrational, in different respects. Second, we can try to find a principled way to exclude aspects like the end of marital harmony from the rational fold. I am perhaps more open to contradictions than the average philosopher outside of Australia, but here I'll explore how to find a principled way to maintain univocal rational assessment.

Though I haven't stressed the affinity, the account on offer here is deeply simpatico with a virtue account of goodness. Norms on this picture are those standards that pick out good-making features of the object of the norm. So, I suggest we employ a unity of the virtues account here to see if we can get traction to naturally separate the functions in question into parts. If we think it is obviously true of the rational (or any topical) mode of assessment that it must be univocal (for every question, it must not be the case that there are norms that speak in its favor and against it), we should be sympathetic to a version of the unity of the virtues thesis anyway: it should be possible to satisfy one norm without thereby violating another of the same sort. Of course we saw that this cannot be, in general, true. It often takes a normative sacrifice to achieve local greatness. But within one domain we desire the possibility of perfection, or at least that it not be hindered by the very normative practice set to evaluate it.

If this is so, then it should be possible, whenever you are incapable of ideally exercising two functions simultaneously, that they are governed by norms of different types. So, it was a mistake to originally pick out theoretical norms as just those that govern mental capacities while determining that rationality was the property of those mental functions when they ide-

ally functioned. Because some mental capacities have functions that demand incompatible things. In this case, the principle should be that the functions that can be ideally exercised without violation of the norms of other mental functions are those that comprise theoretical rationality. Because the spousal over-rating and general reasoning functions are not simultaneously ideally exercised, then at least one of them cannot be a criterion of theoretical rationality. In these, and I suspect all, cases more general functions should be privileged while more specific ones are weighed against the general ones.

So, if this is plausible, it seems that we should take the general reasoning function to be the determinant of theoretical rationality-determining norms, while the spousal-ranking function has some other, perhaps socially relevant normative appraisal. It should not be the case, then, that either way one is doomed to irrationality (while also being guaranteed some form of rationality). Instead, one is doomed to do something wrong (either rationally or socially) and something right. While this might strike one as pessimistic, it seems like a generally familiar plight to have.

BIBLIOGRAPHY

Akins, K. (1996) "Of Sensory Systems and the 'Aboutness' of Mental States," *The Journal of Philosophy.* 93:7, 337–372.

Alike, M. (1985) "Global Self-evaluation as Determined by the Desirability and Controllability of Trait Adjectives," *Journal of Personality and Social Psychology.* 49:6, 1621–1630.

Allen, C and M. Bekoff (1995) "Biological Function, Adaptation, and Natural Design," *Philosophy of Science.* 62, 609–22.

Amundson, R and G.V. Lauder (1994) "Function Without Purpose: The Use of Causal Role Function in Evolutionary Biology," *Biology and Philosophy.* 9, 443–69.

Antony, L. (1996) "Equal Rights for Swamp-persons," *Mind & Language* 11:1, 70–75.

Ashwell, L. (2010) "Superficial Dispositionalism," *Journal of Philosophy* 88:4, 635–53.

Barandalla, A. and M. Ridge (2011) "Function and Self-Constitution," *Analysis* 71:2, 364–80.

Bechtel, W. (1986) "Teleological Functional Analyses and the Hierarchical Organization of Nature," In N. Rescher ed. *Current Issues in Teleology.* Langham, MD: University Press of America.

Bechtel, W. (1989) "Functional analyses and their justification," *Philosophy and Biology.* 4:2, 159–162.

Becker, M. (1969) "Function and Teleology," *Journal of the History of Biology.* 2, 151–64.

Bedau, M. (1992) "Where's the Good in Teleology?" *Philosophy and Phenomenological Research,* 52, 781–805.

Bigelow, J and R. Pargetter. (1987) "Functions," *The Journal of Philosophy.* 84, 181-96.

Björnsson, A and Finlay, S (2010) "Metaethical Contextualism Defended," *Ethics* 121: 1, 7–36.

Boorse (2002) "A Rebuttal of Functions," in A. Ariew, R. Cummins, and M. Perlman Eds. in *Functions: New Essays in the Philosophy of Biology,* Oxford: OUP, 63 - 112.

Bratman, M. (2009) "Intention, Belief, and Instrumental Rationality," in *Reasons for Action* eds. D. Sobel and S. Wall.

Broome, J. (2007) "Wide or Narrow Scope?" *Mind.* 116:462, 359–370.

Brown, W. and C. Moore (2000) "Is prospective altruist-detection an evolved solution to the adaptive problem of subtle cheating in cooperative ventures? Supportive evidence using the Wason selection task," *Evolution and Human Behavior.* 21:1, 25–37.

Buller, D.J. (2002) "Function and Design Revisited," in A. Ariew, R. Cummins, and M. Perlman (eds) *Functions: New Essays in the Philosophy of Psychology and Biology.* Oxford: Oxford University Press. 222–43.

Carlisle, E., and Shafir, E. (2005) "Questioning the Cheater-detection hypothesis," *Thinking & Reasoning.* 11:2, 971-122.

Carrara, M. and P. Vermaas. (2009) "The Fine-Grained Metaphysics of Artefactual and Biological Functional Kinds," *Synthese,* 169, 125–143.

Casscells W, Schoenberger A, Graboys TB. (1978) "Interpretation by Physicians of Clinical Laboratory Results," *New England Journal of Medicine.* 299:18, 999–1001.

Chang, R. (2009) "Voluntarist reasons and the sources of normativity," in David Sobel &Steven Wall (eds.), *Reasons for Action.* Cambridge University Press

Chang, R. (2013). "Grounding Practical Normativity: Going Hybrid," *Philosophical Studies* 164:1,163–187.

Choi, S. (2005) "Do Categorical Ascriptions Entail Counterfactual Conditionals?" *Philosophical Quarterly* 55: 495–503.

Clark, P. (2001) "Velleman's Autonomism," *Ethics* 111:3, 580–93

Clarke, R. (2010) "Opposing Powers," *Philosophical Studies* 149, 153–160.

Cosmides, L. and J. Tooby (1992) "Cognitive Adaptations for Social Exchange," in *The Adapted Mind.* Oxford: Oxford University Press.

Cosmides, L. and J. Tooby (1996) "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty," *Cognition.* 58:1, 1–73.

Cummins, R. (1975) "Functional Analysis," *The Journal of Philosophy.* 72, 741–65.

Darwell, S. (1990) "Autonomist Internalism and the Justification of Morals," *Noûs* 24, 257–68.

Davidson, D. (1987) "Knowing Ones Own Mind," *Proceedings and Addresses of the American Philosophical Association.* 60:3, 441–458.

Davies, P.S. (2000) "Malfunctions," *Biology and Philosophy.* 15, 19–38.

Davies, P.S. (2000) "The Nature of Natural Norms: Why Selected Functions are Systemic Capacity Functions," *Noûs* 34, 85–107.

Dennett, D. "Cow-sharks, Magnets, and Swampman," *Mind & Language.* 11:1, 76–77.

Dreier, J. (1997) "Humean Doubts and the Practical Justification of Morality," in *Ethics and Practical Reason*, eds. Cullity and Gaut. Oxford: Clarendon Press.

Dreier, J. (2001) "Humean Doubts About Categorical Imperatives," in *Varieties of Practical Reasoning*, ed. Millgram. Cambridge, MA: MIT Press.

Dretske, F. (1996) "Absent Qualia," *Mind & Language* 11:1, 78–85.

Elder, C (1995) "A Different Kind of Natural Kind," *Australasian Journal of Philosophy.* 73:4, 516–531.

Enç, B. (2002) "Indeterminancy of Function Attribution," in A. Ariew, R. Cummins, and M. Perlman Eds. in *Functions: New Essays in the Philosophy of Biology,* Oxford: OUP.

Enoch, D. (2006) "Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Agency," *Philosophical Review* 115, 169–98.

Enoch, D. (2011) *Taking Morality Seriously.* Oxford: Oxford University Press.

Everett, A. (2009) "Intrinsic Finks, Masks, and Mimics," *Erkenntnis* 71: 191–203.

Fara, M. (2005) "Dispositions and Habituals," *Noûs* 38, 159–74.

Ferrero, L. (2009) "Constitutivism and the Inescapability of Agency," *Oxford Studies in Metaethics* 4, 303–33.

Fiedler, K. (1988) "The dependence of the conjunction fallacy on subtle linguistic factors," *Psychological Research.* 50:2, 123–129.

Fine, K. (1994) "Essence and Modality," *Philosophical Perspectives,* 8, 1–16.

Finlay, S. (2004) "The Conversational Practicality of Value Judgment," *The Journal of Ethics* 8, 205–23.

Finlay, S. (forthcoming) *Confusion of Tongues: A Theory of Normative Language.* Oxford: Oxford University Press.

Foot, p. (2001) *Natural Goodness.* Oxford: Oxford University Press.

Foot, P. (2002) *Virtues and Vices; second edition.* Oxford: Oxford University Press.

Frankish, K. (2010) "Dual-Process and Dual-System Theories of Reasoning," *Philosophy Compass.* 5:10, 914–926.

Franssen, M. (2006) "The Normativity of Artifacts," *Studies in the History and Philosophy of Science.* 44, 409–22.

Gagné, F. and J. Lydon (2004) "Bias and Accuracy in Close Relationships: An Integrative Review," *Personality and Social Psychology Review.* 8:4, 322–338.

Garson, J. (2011) "Selected Effects Functions and Causal Role Functions in the Brain: The Case for an Etiological Approach to Neuroscience," *Philosophy & Biology.* 26: 547–565.

Geach, P. (1956) "Good and Evil," *Analysis.* 17, 33–42.

Gigerenzer, G. (1994) "Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa)," in G. Wright &P. Ayton (Eds.) *Subjective Probability.* Wiley, 129–161.

Gigerenzer, G. and P. Todd. (1999) "Fast and frugal heuristics: The adaptive toolbox," in *Evolution and Cognition.* Oxford: Oxford University Press. 3–34.

Godfrey-Smith, P. (1994) "A Modern History Theory of Functions," *Noûs* 28, 344–62.

Griffiths, P.E. (1993) "Functional Analysis and Proper Functions," *British Journal for the Philosophy of Science.* 44, 409–22.

Griggs, J. and R. Cox. (1982) "The effects of experience on performance in Wason?s selection task," *Memory & Cognition.* 10:5, 496–502.

Hàjek, A. "Interpretations of Probability," *The Stanford Encyclopedia of Philosophy.* (Winter 2012 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/win2012/entries/probability-interpret/.

Horan, B. (1989) "Functional Explanations in Sociobiology," *Biology and Philosophy.* 4:2, 131–158.

Hursthouse, R. (1991) "Arational Actions," *Journal of Philosophy.* 88, 57–68.

Hussain, N. (2004) "The Return of Moral Fictionalism," *Philosophical Perspectives.* 18:1, 149–188.

Jespersen, B. and Carrara, M. (2013) "A New Logic of Technical Malfunction," *Studia Logica.* 101:3, 547–581.

Johnston, M. (1992) "How to Speak of the Colors," *Philosophical Studies* 68, 221–63.

Kitcher, P. (1993) "Function and Design," *Midwest Studies in Philosophy.* 18, 379–97.

Koehler, J.J. (1996) "The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges," *Behavioral and Brain Sciences.* 19:01, 1–17.

Kolodny, N. (2007) "State or Process Requirements?" *Mind.* 116:462, 371–385.

Korsgaard, C. (2009) *Self-Constitution,* Oxford, Oxford University Press.

Korsgaard, C.M. (1996) *The Sources of Normativity.* Cambridge: Cambridge University Press.

Korsgaard, C.M. (2003) "Realism and Constructivism in Twentieth-Century Moral Philosophy," *Journal of Philosophical Research*, APA Centennial Supplement (2003): 99–122. Reprinted in Christine M. Korsgaard, *The constitution of Agency: Essays on Practical Reason and Moral Psychology.* Oxford: Oxford University Press, 2008.

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

Lavin, D. (2004) "Practical Reason and the Possibility of Error," *Ethics* 114, 424–57.

Lichtenstein, S, B. Fischhoff, L.D. Phillips (1982) "Calibration of probabilities: The state of the art," in D. Kahneman, P. Slovic, A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge Univ. Press, Cambridge.

Ludwig, K. (1996) "Duplicating Thoughts," *Mind & Language.* 11:1, 92–102.

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong.* New York: Penguin.

McKay, R. and Dennett, D. (2009) "The Evolution of Misbelief," *Behavioral and Brain Sciences.* 32:06, 493–510.

Mele, A. (1989) "Motivational Internalism: The Powers and Limits of Practical Reasoning," *Philosophia* 19, 417–36.

Mele, A. (2004) "Velleman on Action and Agency," *Philosophical Studies.* 121:3, 249–261.

Millikan, R. (1993) "In Defense of Proper Function," in *White Queen Psychology and Other Essays for Alice.* Cambridge, MA: MIT University Press, 13–29.

Millikan, R.G. (2002) "Biofunctions: Two Paradigms," In A. Anew, R. Cummins, &M. Perlman (Eds.), *Functions: New essays in the philosophy of psychology and biology* 113–143.

Millikan, R.G. (1984) *Language, Thought, and Other Biological Categories: New Foundations for Realism.* Cambridge: MIT Press.

Millikan, R.G. (1989) "In Defense of Proper Functions," *Philosophy of Science,* 56, 288–302.

Millikan, R.G. (1993) "Dispositions or Etiologies? A Comment on Bigelow and Pargetter," *Journal of Philosophy* 90: 249–259.

Millikan, R.G. (1999) "Wings, Spoons, Pills, and Quills: A Pluralist Theory of Function," *The Journal of Philosophy,* 96, 191–206.

Millikan, R.G. (1996) "On Swampkinds," *Mind & Language.* 11:1, 103–117.

Mitchell, S. (1993) "Dispositions or Etiologies? A Comment on Bigelow and Pargetter," *The Journal of Philosophy.* 90: 249–59.

Neander, K. (1991) "The Teleological Notion of Function," *Australasian Journal of Philosophy.*69:4, 454–468.

Neander, K. (1991) "Functions as Selected Effects: The Conceptual Analyst's Defense," *Philosophy of Science,* 58, 168–84.

Neander, K. (1995) "Misrepresenting and Malfunctioning," *Philosophical Studies,* 79, 109–41.

Neander, K. (1996) "Swampman Meets Swampcow," *Mind & Language.* 11:1, 118–129.

Neander, K (2002) "Types of Traits," In Andr Ariew (ed.), *Functions.* Oxford University Press. 402–422.

Perlman, M. (2004) "The Modern Philosophical Resurrection of Teleology," *The Monist,* 87, 3–51.

Preston, B. (2009). "Philosophical Theories of Artifact Function," *Philosophy of technology and engineering sciences.* 9, 213–233.

Preston, B. (1998) "Why is a Wing Like a Spoon? A Pluralist Theory of Function," *The Journal of Philosophy,* 95, 215–54.

Prior, E.W. (1985) "What is Wrong with Etiological Accounts of Biological Function?," *Pacific Philosophical Quarterly,* 66, 310–28.

Railton, P. (1997) "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," In *Ethics and Practical Reason.* eds. Garrett and Cullity, Oxford: Clarendon Press, 53–79.

Raz, J. (2005) "Instrumental Rationality: A Reprise," *Journal of Ethics* & *Social Philosophy* Symposium I, Dec 2005.

Raz, J. (2005) "The Myth of Instrumental Rationality," *Journal of Ethics* & *Social Philosophy* 1, 2–28.

Reisner, A. (2009) "Unifying the Requirements of Rationality," *Philosophical Explorations.* 12:3, 243–260.

Reed, G. M., Kemeny, M. E., Taylor, S. E. &Visscher, B. R. (1999) "Negative HIV-specific expectancies and AIDS-related bereavement as predictors of symptom onset in asymptomatic HIV-positive gay men," *Health Psychology.* 18, 354–63.

Rosati, C. (1995) "Naturalism, Normativity, and the Open Question Argument," *Noûs* 29, 46–70.

Rosati, C. (2003) "Agency and the Open Question Argument," *Ethics* 113, 490–527.

Ross, W.D. (1930) *The Right and The Good.* Oxford: Oxford University Press.

Schroeder, M. (2007) "Teleology, Agent-Relative Value, and 'Good'," *Ethics* 116, 265–95.

Searle, J. (1969) *Speech Acts: An Essay in the Philosophy of Language.* Cambridge, Cambridge University Press.

Setiya, K. (2003) "Explaining Action," *Philosophical Review* 112:3, 339–93.

Setiya, K. (2007) *Reasons Without Rationalism.* Princeton, Princeton University Press.

Setiya, K. (2010) "Sympathy for the Devil," In S. Tenenbaum, ed.*Desire, Practical Reason, and the Good.* Oxford: Oxford University Press, 82–110.

Shah, N. (2003) "How Truth Governs Belief," *Philosophical Review* .112:4, 447–482.

Sober, E. (1993) *Philosophy of Biology,* Boulder: Westview Press.

Sperber, D., Cara, F., & Girotto, V. (1995) "Relevance Theory Explains the Selection Task," *Cognition.* 57:1, 31–95.

Sperber, D., & Girotto, V. (2002). "Does the selection task detect cheater-detection?" In J. Fitness & K. Sterelny (Eds.), *New directions in evolutionary psychology.* Hove, UK: Psychology Press.

Taylor, S. E., Kemeny, M. E., Reed, G. M., Bower, J. E. &Gruenewald, T. L. (2000) "Psychological resources, positive illusions, and health," *American Psychologist.* 55, 99–109.

Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M. &McDowell, N. K. (2003) "Are self-enhancing cognitions associated with healthy or unhealthy biological profiles?" *Journal of Personality and Social Psychology.* 85:4, 605–15.

Thomson, J.J. (1992) "On Some Ways in which a Thing Can Be Good," *Social Philosophy and Policy* 2, 96–117.

Thomson, J.J. (1997) "The Right and the Good," *Journal of Philosophy* 94, 273–98.

Thomson, J.J. (2008) *Normativity.* Chicago, Open Court Press.

Tiffany, E. (2012). "Why be an Agent?" *Australasian Journal of Philosophy.* 90:2, 223–233.

Tinbergen, N. (1963) "On Aims and Methods of Ethology," *Zeitschrift für Tierpsychologie* 20: 410–429.

Trivers, R. L. (1971) "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology.* 46, 35–57.

Tversky, A., & Kahneman, D. (1983) "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment," *Psychological Review.* 90:4, 293-315.

Velleman, J.D. (1996) "The Possibility of Practical Reason," reprinted in his *The Possibility of Practical Reason*. Oxford: Oxford University Press, 2000, 170–99.

Velleman, J.D. (2000) *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Velleman, J.D. (2004) "Replies to the Discussion on *The Possibility of Practical Reason*," *Philosophical Studies* 121, 277–98.

Velleman, J.D. (2009) *How We Get Along.* Cambridge: Cambridge University Press.

Vermass, P.E. and W. Houkes (2003) "Ascribing Functions to Technical Artefacts: A Challenge to Etiological Accounts of Functions," *British Journal for the Philosophy of Science,* 54, 261–89.

Walsh, D. M. (1996) "Fitness and Function," *The British Journal for the Philosophy of Science.* 47:4, 553–574.

Walsh, D.M. and A. Ariew (1996) "A Taxonomy of Functions," *Canadian Journal of Philosophy,* 26, 493–514.

Wenger, A. & Fowers, B. J. (2008) "Positive illusions in parenting: Every child is above average," *Journal of Applied Social Psychology.* 38:3, 611–34.

Wright, L. (1973) "Functions," *Philosophical Review,* 82, 139–68.

Ziff, P. (1960) *Semantic Analysis*. Ithaca: Cornell University Press.